LIMIT THEOREMS IN FUNCTIONAL DATA ANALYSIS WITH APPLICATIONS

by

Ron Reeder

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics

The University of Utah

December 2011

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Ron Reeder**

has been approved by the following supervisory committee members:

Lajos Horvath	, Chair	**30 Sep 2011** Date Approved
Davar Khoshnevisan	, Member	**30 Sep 2011** Date Approved
Firas Rassoul-Agha	, Member	**30 Sep 2011** Date Approved
David Kiefer	, Member	**30 Sep 2011** Date Approved
Andrejs Treibergs	, Member	**30 Sep 2011** Date Approved

and by **Peter Trapa** , Chair of

the Department of **Mathematics**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

This dissertation is concerned with functional data analysis. Functional data consists of a collection of curves or functions defined on an interval. These curves can be obtained by splitting a continuous time record such as temperature into daily or annual curves. Functional data is also obtained when an experimenter records a curve of data from each subject in a sample, e.g., a growth trajectory of an animal or plant. Several examples of different models for functional data are given. We use the method of principle component analysis to obtain the necessary regularization in each model. Functional principal component analysis is summarized as a natural extension of the traditional vector principal component analysis.

The first functional model is concerned with inference based on the mean function of a functional time series. We develop and asymptotically justify a testing procedure for the equality of means in two functional samples exhibiting temporal dependence.

As a second example, we consider a quadratic functional regression model in which a scalar response depends on a functional predictor. We develop a test of the significance of the nonlinear term in the model. The asymptotic behavior of our testing procedure is established.

In the third model, we observe two sequences of curves which are connected via an integral operator. This model includes linear models as well as autoregressive models in Hilbert spaces. We develop a procedure to test the stability of the model.

In the fourth model, we propose a functional version of the popular ARCH model. We establish conditions for the existence of a strictly stationary solution, derive weak dependence and moment conditions, show consistency of the estimators, and perform an empirical study demonstrating how our model matches with real data.

For my family.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to say a few words about those who have guided and supported me. My family, especially my wife, Andrea, has given me constant encouragement. I also appreciate the professors who have taught and inspired me.

I was not sure that I would pursue a career in math when I met Christopher Grant, a professor at Brigham Young University. I was taking a first course in analysis, hoping that I would get a better idea of what it would be like to be a mathematician. It was in this class that I learned the value of a carefully crafted argument. Dr. Grant seemed to have an "open door" policy. I don't recall if he actually had official office hours, but he was nearly always in his office with the door open and willing to help at nearly any time. When I first began going to his office for help, I asked questions like, "How do you do problem 5." His response was a dramatic pause, followed by a question. Another long pause would follow and then another question. He listened very carefully to everything I said so that he could see what I understood and what I misunderstood. I soon learned that in these long pauses he was considering what question he could ask that would force me to think very carefully about the obstacle in my path. Thank you, Dr. Grant. This was the most influential class of my entire education. I agree with my classmate, Ian, who said, "It changes the way you think."

I met Davar Khoshnevisan during my first semester as a graduate student at the University of Utah. Davar was teaching the basic probability course and his careful arguments and attention to detail reminded me of Dr. Grant. Even when we were just talking about dice and cards, he made probability seem intrinsically interesting. I was interested in probability and I liked Davar's teaching style so much that I decided that whenever I had an opportunity to take another class from him, I would. In the end, I have taken more classes from Davar than any other professor. Thank you, Davar. You have had a profound effect on the academic path that I have chosen.

Lajos Horváth is my advisor and has been my greatest mentor. I had been a student at the University of Utah for almost two years when I met Lajos. I was interested in taking our statistics sequence, but my schedule did not allow it. So Masaki and I went to Lajos, whom we did not know, to ask if we could take statistics as a reading course. He was surprised

that two PhD students who were interested in statistics had somehow gone unnoticed for two years. He suggested that if we were interested in statistics, then we could work with him and should begin right away. I accepted his invitation. I was surprised that Lajos was willing to work with me since I did not have the background in statistics. He seemed to believe that the best way to learn is by doing. He immediately started me on a project to find a necessary and sufficient condition for the existence of a solution to a certain time series model. We would meet together in the lounge and he would explain things at the blackboard that I needed to know. I was again surprised by the amount of time he was willing to spend helping me. We have written several papers together since then. Under his tutelage, I have improved my writing and presenting skills and, of course, learned a great deal about statistics. Lajos has encouraged me and supported me all along the way. I could not have picked a better advisor. Thank you, Lajos. You have given me a world of opportunities.

CHAPTER 1

INTRODUCTION

In this chapter, we introduce functional data and discuss a few examples. Then, we give a brief outline of this dissertation.

1.1 Examples of functional data

One of the main goals of statistics is to make inferences about relationships or parameters based on the data that we collect. These data may take several different forms. A sample of data for a stock price like Microsoft may consist of a time series of closing prices. A bank may wish to use this time series to determine an appropriate or fair price for a call option on Microsoft stock. Data may also come in the form of vectors. For example, a doctor might record the height, weight, age, and blood pressure of each patient. In this situation, each observation is a vector of length 4, where the first entry is height, the second is age, etc. A doctor may use this or similar types of data to determine a patient's risk of developing diabetes or other diseases.

The Microsoft stock price data could also be recorded as vectors. If we measure the value of a Microsoft stock at 1-minute intervals throughout the trading day, we obtain a vector of length 390 for each day. The value of a stock is actually known exactly at the times when it is traded. Some stocks trade several times per second, whereas other stocks trade less than 10 times in a day. The times of the day at which a particular stock is traded will be different each day, and the number of times it is traded may also vary from day to day. If we are using a vector to represent each day, then the size of the vector will change from day to day. Also, elements in the same location in the vector from two different days may not correspond to the same time of day. One approach is to use a curve or a function to represent the values of the stock throughout a trading day. Since the value of the stock is only known at times when it is traded, we obtain a discrete record. However, this discrete data can be thought of as an approximation to a continuous time record. The continuous time record is then approximated by some method of interpolation such as cubic splines. A similar data set is the S&P 100 index, which is pictured in Figure 1.1.

The temperature at an observatory could similarly be represented by curves or functions. It is true that the temperature is measured only at a finite number of times throughout the day. However, unlike stock prices, those measurements really are an approximation to a continuous time record. The temperature is defined at any time throughout the day, even if it is not measured.

The techniques that we will develop for the analysis of this type of functional data requires that the data be segmented. In other word, we don't use a single function to represent all of our data. For Microsoft stock, it makes sense to segment the data into days because there are several hours separating one trading day from the next. In Figure 1.1, vertical lines separate the trading days. In contrast, there is no gap in measurement of temperature. However, one of the most influential factors in determining temperature is the daily rotation of the earth. This contributes a very strong daily periodic element to the data, suggesting segmentation into days. Depending on the purpose of the investigation, it is also reasonable to segment temperatures into years.

As another example of functional data, consider the egg-laying trajectories of female Mediterranean fruit flies. We first obtain discrete data by counting the number of eggs that a fly lays on each day of its life. Then, we interpolate this data to obtain a continuous egg-laying curve for each fly. Ten such curves are pictured in Figure 1.2. One difference in this example is that egg-laying curves are likely to be independent, whereas stock prices on Friday are certainly not independent of the price on Thursday.

We have seen in the examples of Microsoft stock, temperature, and fruit flies, that it often makes sense to record data as sequences of curves or functions. This type of data is called functional data and is the focus of this dissertation. In the remainder of this work, functional data will be of the form $\{X_n(t), n = 1, 2, \ldots,\ N,\ t \in [a, b]\}$. By shifting and scaling, we can assume that these functions are defined on the interval $[0, 1]$ so that functional data is of the form $\{X_n(t),\ n = 1, 2, \ldots, N,\ t \in [0, 1]\}$. In the stock (temperature) example, $X_n(t)$ is the value of the stock (temperature) on day n at time t. For the fruit fly example, $X_n(t)$ is the number of eggs laid by the n^{th} fly on day t.

1.2 Organization of the dissertation

This dissertation focuses on modeling and analysis with functional data. It consists of 6 chapters. Chapter 1 describes the type of data we are concerned with and gives an outline of the remaining chapters. In Chapter 2 some general methods and techniques for functional data analysis are introduced. Chapters 3–6 are each devoted to the development and/or

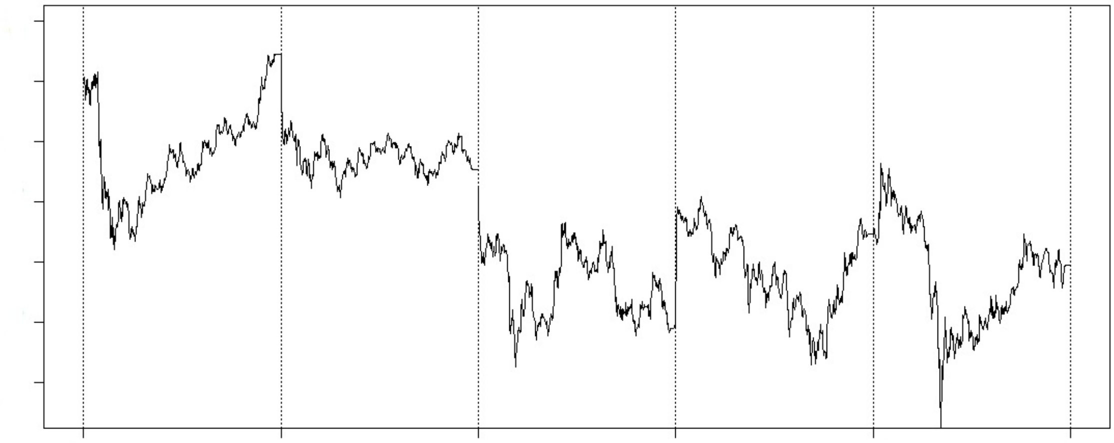


Figure 1.1. S&P 100 index, March 26 to March 30,2007

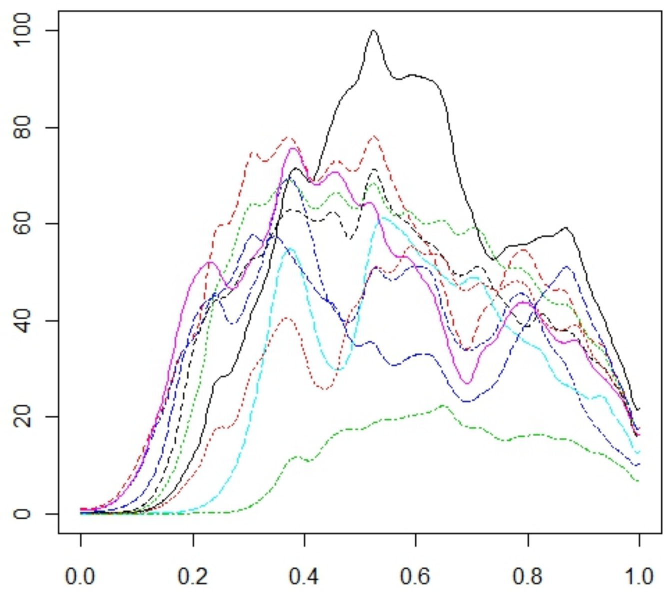


Figure 1.2. Ten smoothed egg-laying curves of Mediterranean fruit flies.

inferences for a specific functional model.

CHAPTER 2

FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS

One of the main goals of principal component analysis (PCA) is a reduction in the dimension of a data set. This is especially helpful when the observations have a very high dimensionality but the intrinsic dimension of the data is much lower. In this chapter, we will assume that vector-valued observations, $\{\mathbf{X}_n, n = 1, \ldots, N\}$, are iid and that the covariance matrix exists. Regarding the function-valued observations, $\{X_n(t), n = 1, \ldots, N\}$, we will assume in this chapter that these are iid and that $E\left(\int_0^1 (X_n(t))^2\, dt\right)^2 < \infty$. Generalizations to this case will be made as necessary in the sequel.

2.1 Vector PCA

In the case when observations are vectors of length L, one can often explain the variance-covariance structure of these variables with only a few linear combinations of these vectors. These linear combinations are referred to as principal components. Thus if we use only p principal components, we have reduced the dimensionality of the observations from L to p. Essentially, we are projecting our observations onto a lower dimensional subspace to reduce the dimensionality. The subspace that we use is chosen so that we retain as much information (variation) in the data set as possible. The principal components are the p uncorrelated random variables, $\{Y^{(i)} = \langle \mathbf{X}, \mathbf{v}_i \rangle, i = 1, \ldots, p\}$ such that $\operatorname{Var}\left(\sum_{i=1}^{p} \langle \mathbf{X}, \mathbf{v}_i \rangle\right)$ is maximized subject to the constraint that $\|\mathbf{v}_i\| = 1$. It is shown in Johnson and Wichern (2007) that the vectors, $\mathbf{v}_i$, that maximize this variance are the eigenvectors of the covariance matrix, $\mathbf{\Sigma} = \operatorname{Var}(\mathbf{X})$, associated with the largest eigenvalues.

In practice, however, the covariance matrix, $\mathbf{\Sigma}$, is not known and must be estimated from the sample. The standard estimators for $\mathbf{\Sigma}$ are

$$\hat{\mathbf{\Sigma}} = \frac{1}{N} \sum_{n=1}^{N} \left(\mathbf{X}_n - \bar{\mathbf{X}}\right)\left(\mathbf{X}_n - \bar{\mathbf{X}}\right)^T,$$

which is the maximum likelihood estimator if $\mathbf{X}_n$ is normally distributed, and

$$\mathbf{S} = \frac{1}{N-1}\sum_{n=1}^{N}\left(\mathbf{X}_n - \bar{\mathbf{X}}\right)\left(\mathbf{X}_n - \bar{\mathbf{X}}\right)^T.$$

Either choice is very reasonable and the difference between the two is small since N is usually rather large. We will use $\hat{\mathbf{\Sigma}}$. Now that we have an estimate for $\mathbf{\Sigma}$, we can find the eigenvectors and eigenvalues of $\hat{\mathbf{\Sigma}}$ and use these as approximations to the eigenvalues of $\mathbf{\Sigma}$ in order to perform principal component analysis. Let $\hat{\mathbf{v}}_i$ ($\mathbf{v}_i$) be the eigenvector associated with the i^{th} largest eigenvalue, $\hat{\lambda}_i$ (λ_i), of $\hat{\mathbf{\Sigma}}$ ($\mathbf{\Sigma}$). We would like to have $\hat{\mathbf{v}}_i$ be close to $\mathbf{v}_i$. However, distinct eigenvalues do not determine the sign of the eigenvector. Therefore, the most we can expect is to have $\hat{c}_i\hat{\mathbf{v}}_i$ be close to $\mathbf{v}_i$, where $\hat{c}_i = \text{sign}(\langle\hat{\mathbf{v}}_i, \mathbf{v}_i\rangle)$. Principal component analysis projects observations onto the space spanned by the first p eigenvectors. The possible sign discrepancy in the estimate does not harm us because the sign of the basis vectors have no effect on the space that is spanned.

2.2 Functional PCA

The samples of curves, $X_n(t)$, that we considered in Section 1.1 are viewed as the outcomes of random variables. These random variables are also denoted by $X_n(t)$. As in the case of vector-valued observations, our goal is to reduce the dimensionality of the observations by projecting onto a lower dimensional subspace. In the case where observations are functions, this is even more significant because we project from an infinite dimensional space, $L^2[0,1]$, to a finite dimensional space.

The subspace that we use is chosen so that we retain as much information (variation) in the data set as possible. The principal components are the p uncorrelated random variables, $Y^{(i)} = \langle X(t), v_i(t)\rangle, i = 1, \ldots, p$ such that $\text{Var}\left(\sum_{i=1}^{p}\langle X(t), v_i(t)\rangle\right)$ is maximized subject to the constraint that $\|v_i\| = 1$.

It is shown in Horváth and Kokoszka (2011) that the functions, $v_i(t)$, that maximize this variance are the eigenfunctions associated with the largest eigenvalues of the covariance operator, C. To define the covariance operator, we first define the covariance function of $X_n(t)$ as $c(s,t) = \text{Cov}(X_n(t), X_n(s))$. The covariance operator, C, is then defined by $C(f) = \int_0^1 f(t)c(s,t)\,dt$. The eigenvalues and eigenfunctions of the operator, C, satisfy the equation $\int_0^1 v_i(t)c(s,t)\,dt = \lambda_i v_i(s)$.

In practice, however, the covariance function and hence the covariance operator are not known and must be estimated from the sample. A standard estimator for $c(s,t)$ is

$$\hat{c}(s,t) = \frac{1}{N}\sum_{n=1}^{N}\left(X_n(t) - \bar{X}(t)\right)\left(X_n(s) - \bar{X}(s)\right).$$

Using $\hat{c}(s,t)$ we obtain an estimate, $\hat{C}$, for the covariance operator C. Now that we have an estimate for C, we can find the eigenfunctions and eigenvalues of $\hat{C}$ and use these as approximations to the eigenvalues of C in order to perform principal component analysis. Let $\hat{v}_i(t)$ $(v_i(t))$ be the eigenfunction associated with the i^{th} largest eigenvalue, $\hat{\lambda}_i$ (λ_i), of $\hat{C}$ (C). We would like to have $\hat{v}_i(t)$ be close to $v_i(t)$. However, distinct eigenvalues do not determine the sign of the eigenfunction. Therefore, the most we can expect is to have $\hat{c}_i\hat{v}_i(t)$ be close to $v_i(t)$, where $\hat{c}_i = \text{sign}(\langle \hat{v}_i(t), v_i(t)\rangle)$. Principal component analysis projects observations onto the space spanned by the first p eigenfunctions. The possible sign discrepancy in the estimate does not harm us because changing the signs of basis functions has no effect on the space they span.

The following important approximation using principal components will be used throughout this dissertation:

$$\begin{aligned} X_n(t) &= \sum_{i=1}^{\infty} \langle X_n(s), v_i(s)\rangle v_i(t) \\ &\approx \sum_{i=1}^{p} \langle X_n(s), v_i(s)\rangle v_i(t) \\ &\approx \sum_{i=1}^{p} \langle X_n(s), \hat{c}_i\hat{v}_i(s)\rangle v_i(t). \end{aligned}$$

2.3 Choosing the number of principal components

We have seen that principal components can be used to reduce the dimensionality of vector- or function-valued observations. The number of principal components that we use is an important decision. We would like to reduce the dimensionality as much as we reasonably can. However, if data are intrinsically five-dimensional and we project onto a two-dimensional subspace, we lose too much information. There are many methods of choosing the appropriate number of principal components to use, including pseudo AIC, cross-validation, or using a Scree plot. In this dissertation, we use the method prescribed by Ramsay and Silverman (2005), which is to choose p so that approximately 85% of the variance in the data is explained by the first p principal components. The proportion of the variance explained by the first p principal components is given by the CPV function. CPV stands for cumulative percentage of total variance and is defined by

$$CPV(p) = \frac{\sum_{i=1}^{p} \hat{\lambda}_i}{\sum_{i=1}^{N} \hat{\lambda}_i}.$$

2.4 Bibliography

Horváth, L. and P. Kokoszka (2011+). *Inference for Functional Data with Applications.* Springer Series in Statistics. Springer. Forthcoming.

Johnson, R. and D. Wichern (2007). *Applied Multivariate Statistical Analysis.* Prentice Hall.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis.* Springer.

CHAPTER 3

ESTIMATION OF THE MEAN OF FUNCTIONAL TIME SERIES AND A TWO SAMPLE PROBLEM[1]

This chapter is concerned with inference based on the mean function of a functional time series, which is defined as a collection of curves obtained by splitting a continuous time record, e.g. into daily or annual curves. We develop a normal approximation for the functional sample mean, and then focus on the estimation of the asymptotic variance kernel. Using these results, we develop and asymptotically justify a testing procedure for the equality of means in two functional samples exhibiting temporal dependence. Evaluated by means of a simulations study and application to real data sets, this two sample procedure enjoys good size and power in finite samples. We provide the details of its numerical implementation.

3.1 Introduction

Functional time series form a class of data structures which occurs in many applications, but several important aspects of estimation and testing for such data have not received as much attention as for functional data derived from randomized experiments. In the latter case, the curves can often be assumed to form a simple random sample, in particular, the functional observations are independent. For curves obtained from splitting a continuous (in principle) time records into, say, daily or annual curves, the assumption of independence is often violated. This chapter focuses on the methodology and theory for the estimation of the mean function of a functional time series, and on inference for the mean of two functional time series. Despite their central importance, these issues have not yet been studied. The contribution of this chapter is thus two-fold: 1) we develop a methodology and an asymptotic

[1]The content of this chapter is based on joint research with Lajos Horváth and Piotr Kokoszka. It has been submitted to the Journal of the Royal Statistical Society: Series B.

theory for the estimation of the variance of the sample mean of temporally dependent curves under model-free assumptions; 2) we propose procedures for testing equality of two mean functions in functional samples exhibiting temporal dependence.

A functional time series $\{X_k,\, k \in \mathbb{Z}\}$ is a sequence of curves $X_k(t),\ t \in [a,b]$. After normalizing to the unit interval, the curves are typically defined as $X_k(t) = X(k+t),\ 0 \le t \le 1$, where $\{X(u), u \in \mathbb{R}\}$ is a continuous time record, which is often observed at equispaced dense discrete points. An example is given in Figure 3.1, which shows seven consecutive functional observations. More examples are studied in Horváth and Kokoszka (2011). A central issue in the analysis of such data is to take into account the temporal dependence of the observations. The monograph of Bosq (2000) studies the theory of linear functional time series, focusing on the functional autoregressive model. For many functional time series it is however not clear what specific model they follow, and for many statistical procedures it is not necessary to assume a specific model. In this chapter, we assume that the functional time series is stationary, but we do not impose any specific model on it. We assume that the curves are dependent in a very broad sense, which is made precise in Section 3.1.1. The dependence condition we use is however satisfied by all models for functional time series used to date, including the linear, multiplicative, bilinear and ARCH type processes. We refer to Hörmann and Kokoszka (2010) and Aue et al. (2011) for examples.

A direct motivation for the research presented in this chapter comes from a two sample problem in which we wish to tests if the mean functions of two functional time series are equal. A specific problem, studied in greater depth in Section 3.4, is to test if the mean curves of certain financial assets are equal over certain periods. This in turn allows us to conclude whether the expectations of future market conditions are the same or different at specific time periods. In general, if the same mean is assumed for the whole time series, whereas, in fact, it is different for disjoint segments, the inference or exploratory analysis that follows will be faulty, as all prediction and model fitting procedures for functional time series start with subtracting the sample mean, viewed as an estimate of the unique population mean function. The same holds true for independent curves; if two subsamples have different mean functions, subtracting the sample mean function based on the whole data set will lead to spurious results. Despite the importance of two sample problems for functional data, they have received little attention. Recent papers of Horváth et al. (2009) and Panaretos et al. (2010) are the only contributions to a two sample problem in a functional setting which develop inferential methodology. Horváth et al. (2009) compare linear operators in two functional regression models. Panaretos

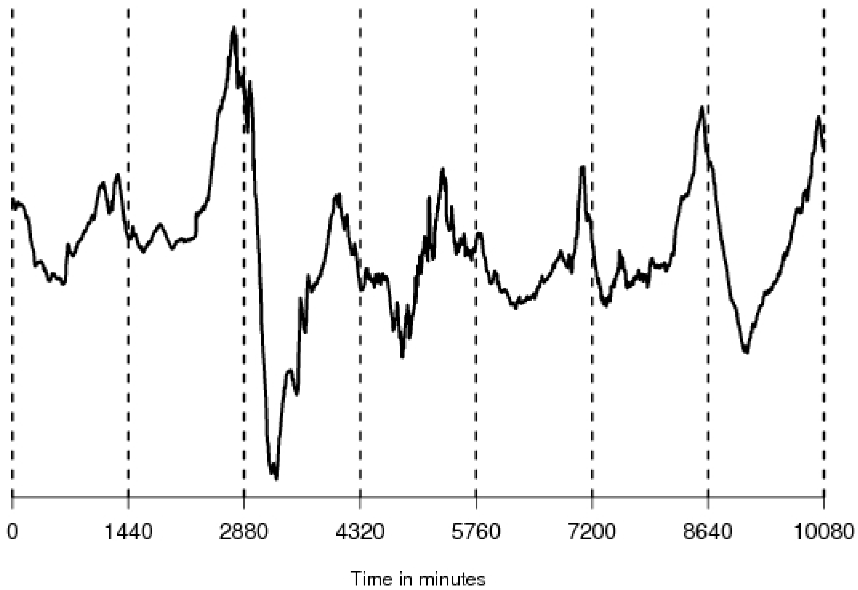


Figure 3.1. The horizontal component of the magnetic field measured in one minute resolution at Honolulu magnetic observatory from 1/1/2001 00:00 UT to 1/7/2001 24:00 UT.

et al. (2010) focus on testing the equality of the covariance operators in two samples of iid Gaussian functional observations; this chapter focuses on the means of dependent (and non-Gaussian) observations. We develop the required methodology, justify it by asymptotic arguments, and describe its practical implementation.

Any inference involving mean functions requires estimates of the variability of the sample mean. In iid functional samples, the sample covariance operator is used, but for functional time series this problem is much more difficult. For scalar and vector-valued time series, the variance of the sample mean is asymptotically approximated by the long-run variance whose estimation has been one of the central problems of time series analysis, studied in textbooks, see e.g. Anderson (1971), Brockwell and Davis (1991), Hamilton (1994), and dozens of influential papers, see Newey and West (1987), Andrews (1991) and Andrews and Monahan (1992), to name just a few. Convergence of various estimators of the long-run variance has been established under several types of assumptions, including broad model specifications (e.g. linear processes), cumulant conditions, and various mixing conditions. Hörmann and Kokoszka (2010) and Gabrys et al. (2010) advocated using the notion of L^p-m-approximability for functional time series, as this condition is intuitively appealing and is easy to verify for functional time series models. We therefore develop a general framework for the estimation of the long-run covariance kernel in this setting. The long-run covariance kernel (equivalently operator) is an infinite dimensional object, precise definitions are given in Section 3.2, whose estimation has not been studied yet. Hörmann and Kokoszka (2010) studied the estimation of a long-run covariance matrix obtained by projecting this operator onto a finite dimensional basis. Their approach is based on the results for vector-valued time series, and they use cumulant-like assumptions which are difficult to verify for nonlinear functional time series.

The long-run covariance kernel corresponds to the asymptotic variance in a normal approximation for the sample mean of a scalar time series, but no central limit theorem for a general functional time series has been established yet (results for linear processes are established in Bosq (2000)). We provide such a generally applicable result as well (Theorem 3.1).

The remainder of the chapter is organized as follows. We conclude the introduction by defining the notion of dependence for functional time series which we use throughout this chapter. Then, in Section 3.2, we state the asymptotic results for the mean of a single functional time series, with proofs developed in Section 3.5. Section 3.3 focuses on the problem of testing the equality of means of two functional samples which exhibit temporal dependence. In Section 3.4, we evaluate the finite sample performance of the procedures proposed in Section 3.3 by means of a simulation study and application to real data. The proofs of the theorems stated in Section 3.3 are collected in Section 3.6.

3.1.1 Approximable functional time series

We consider a stationary functional time series $\{\varepsilon_i(t),\ i \in \mathbb{Z},\ t \in [0,1]\}$, which we can view as an error sequence in a more complex functional model, for example a regression model, as in Gabrys et al. (2010). We assume that these errors are nonlinear moving averages $\varepsilon_i = f(\delta_i, \delta_{i-1}, \ldots)$, for some measurable function $f : S^\infty \to L^2$, and iid elements δ_i of a measurable space S. In all models used in practice $S = L^2$. To motivate the construction below, it is useful to write the ε_i as

$$\varepsilon_i = f(\delta_i, \ldots, \delta_{i-m+1}, \delta_{i-m}, \delta_{i-m-1} \ldots). \tag{3.1}$$

Under (3.1), the sequence $\{\varepsilon_i\}$ is stationary and ergodic. The function f must decay sufficiently fast to ensure that the sequence $\{\varepsilon_i\}$ is weakly dependent. The weak dependence condition is stated in terms of an approximation by m-dependent sequences, namely, we require that

$$\sum_{m \geq 1} \left[E \int \left(\varepsilon_i(t) - \varepsilon_{i,m}(t)\right)^2 dt \right]^{1/2} < \infty, \tag{3.2}$$

where

$$\varepsilon_{i,m} = f(\delta_i, \ldots, \delta_{i-m+1}, \delta^{(m)}_{i,i-m}, \delta^{(m)}_{i,i-m-1}, \ldots), \tag{3.3}$$

with the sequences $\{\delta^{(m)}_{i,k}\}$ being independent copies of the sequence $\{\delta_i\}$. Note that the sum in (3.2) does not depend on i.

The idea behind the above construction is that the function f decays so fast that the effect of the innovations δ_i far back in the past becomes negligible; they can be replaced by different, fully independent innovations. If the ε_i follow a linear model $\varepsilon_i = \sum_{j \geq 0} c_j(\delta_{i-j})$, condition (3.2) intuitively means that the approximations by the finite moving averages $\varepsilon_{i,m} = \sum_{0 \leq j \leq m} c_j(\delta_{i-j})$ become increasingly precise. This means that the operators c_j must decay sufficiently fast in an appropriate operator norm. We refer to Hörmann and Kokoszka (2010) and Aue et al. (2011) for the details and examples of nonlinear functional time series satisfying (3.2).

We also note that the general idea of using nonlinear moving averages (Bernoulli shifts) and imposing moment conditions to quantify dependence has been recently used in other contexts, see Wu (2005, 2007). The connections between such notions and the traditional mixing conditions, or other notions of weak dependence, e.g. that introduced by Doukhan and Louhichi (1999), are only partially understood at present. In particular, it is not clear which functional time series models satisfy dependence conditions other than the approximability (3.2).

3.2 Normal approximation and long-run variance for functional time series

In this section, we state the central limit theorem for the sample mean of an L^2-m-approximable functional time series. Its applicability depends on the estimation of the covariance kernel of the limit. We therefore also establish the consistency of the kernel estimator of the long-run covariance kernel. We assume that $\{\varepsilon_i\}$ is an L^2-m-approximable (and hence stationary) functional time series satisfying

$$E\varepsilon_0 = 0, \quad \text{in } L^2 \tag{3.4}$$

and

$$\int E\varepsilon_0^2(t)\, dt < \infty. \tag{3.5}$$

Theorem 3.1. *If (3.1), (3.2), (3.4), (3.5) hold, then*

$$N^{-1/2} \sum_{i=1}^{N} \varepsilon_i \xrightarrow{d} Z \quad in \quad L^2, \tag{3.6}$$

where Z is a Gaussian process with

$$EZ(t) = 0 \quad \text{and} \quad E[Z(t)Z(s)] = c(t,s);$$

$$c(t,s) = E\varepsilon_0(t)\varepsilon_0(s) + \sum_{i\geq 1} E\varepsilon_0(t)\varepsilon_i(s) + \sum_{i\geq 1} E\varepsilon_0(s)\varepsilon_i(t). \tag{3.7}$$

The infinite sums in the definition of the kernel c converge in $L^2([0,1]\times[0,1])$, i.e. c is a square integrable function on the unit square.

Theorem 3.1 is proven in Section 3.5.

The kernel c is defined analogously to the long-run variance of a scalar time series. It is directly related to the covariance operator of the sample mean defined by

$$\begin{aligned}\hat{C}_N(x) &= NE\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i, x\right\rangle \frac{1}{N}\sum_{j=1}^{N}\varepsilon_j\right] \\ &= \frac{1}{N}\sum_{i,j=1}^{N} E\left[\langle \varepsilon_i, x\rangle\, \varepsilon_j\right].\end{aligned}$$

If the ε_i are independent, then $\hat{C}_N(x) = N^{-1}\sum_{i=1}^N E\left[\langle \varepsilon_i, x\rangle\, \varepsilon_i\right]$ becomes the usual sample (empirical) covariance operator, which plays a central role in many exploratory and inferential tools of functional data analysis of iid functional observations, mostly through the empirical functional principal components defined as its eigenfunctions. For functional time series, it is not suitable. For any stationary functional time series $\{\varepsilon_i\}$,

$$\begin{aligned}\hat{C}_N(x)(t) &= \int \left(\frac{1}{N}\sum_{i,j=1}^{N} E[\varepsilon_i(s)\varepsilon_j(t)]\right) x(s)\, ds \\ &= \int c_N(t,s)x(s)\, ds,\end{aligned}$$

where

$$c_N(t,s) = \sum_{|k|<N} \left(1 - \frac{|k|}{N}\right) E[\varepsilon_0(s)\varepsilon_k(t)]. \tag{3.8}$$

The summands in (3.8) converge to those in (3.7), but the estimation of the long-run covariance kernel c is far from trivial.

To enhance the applicability of our result, we state it for the case of a nonzero mean function, which is estimated by the sample mean. We thus assume that

$$X_i(t) = \mu(t) + \varepsilon_i(t), \quad 1 \le i \le N, \tag{3.9}$$

with the series $\{\varepsilon_i\}$ satisfying the assumptions of Theorem 3.1.

Let K be a kernel (weight) function defined on the line and satisfying the following conditions:

$$K(0) = 1, \tag{3.10}$$

$$K \text{ is continuous}, \tag{3.11}$$

$$K \text{ is bounded}, \tag{3.12}$$

$$K(u) = 0, \text{ if } |u| > c, \quad \text{for some } c > 0. \tag{3.13}$$

Condition (3.13) is assumed only to simplify the proofs, a sufficiently fast decay could be assumed instead.

Next we define the empirical (sample) correlation functions

$$\hat{\gamma}_i(t,s) = \frac{1}{N} \sum_{j=i+1}^{N} \left(X_j(t) - \bar{X}_N(t)\right)\left(X_{j-i}(s) - \bar{X}_N(s)\right), \quad 0 \le i \le N-1, \tag{3.14}$$

where

$$\bar{X}_N(t) = \frac{1}{N} \sum_{1 \le i \le N} X_i(t).$$

The estimator for c is given by

$$\hat{c}_N(t,s) = \hat{\gamma}_0(t,s) + \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)\left(\hat{\gamma}_i(t,s) + \hat{\gamma}_i(s,t)\right) \tag{3.15}$$

where $h = h(N)$ is the smoothing bandwidth satisfying

$$h(N) \to \infty, \quad \frac{h(N)}{N} \to 0, \quad \text{as } N \to \infty. \tag{3.16}$$

In addition to (3.2), we also assume that

$$\lim_{m\to\infty} m \left[E \int \left(\varepsilon_n(t) - \varepsilon_{n,m}(t)\right)^2 \, dt\right]^{1/2} = 0. \tag{3.17}$$

Theorem 3.2. *Suppose the functional time series $\{X_i\}$ follows model (3.9). Under conditions (3.1), (3.2), (3.4), (3.5), (3.10)–(3.13), (3.16), (3.17),*

$$\iint \left(\hat{c}_N(t,s) - c(t,s)\right)^2 \, dt\, ds \xrightarrow{P} 0, \tag{3.18}$$

with $c(t,s)$ by defined by (3.7) and $\hat{c}_N(t,s)$ by (3.15).

Theorem 3.2 is proven in Section 3.5. First we use the results of this section in the problem of testing the equality of means in two functional samples.

3.3 Testing the equality of mean functions

We consider two samples of curves, $X_1, X_2, \ldots, X_N$ and $X_1^*, X_2^*, \ldots X_M^*$, satisfying the following location models

$$X_i(t) = \mu(t) + \varepsilon_i(t), \quad X_j^*(t) = \mu^*(t) + \varepsilon_j^*(t). \tag{3.19}$$

The error functions ε_i are assumed to satisfy the conditions stated in Sections 3.1 and 3.2. The functions ε_j^* are assumed to satisfy exactly the same conditions. In particular, their long-run covariance kernel is defined by

$$c^*(t,s) = E\varepsilon_0^*(t)\varepsilon_0^*(s) + \sum_{i\ge1} E\varepsilon_0^*(t)\varepsilon_i^*(s) + \sum_{i\ge1} E\varepsilon_0^*(s)\varepsilon_i^*(t).$$

We assume that

$$\{\varepsilon_i,\ 1 \le i \le N\} \quad \text{and} \quad \left\{\varepsilon_j^*,\ 1 \le j \le M\right\} \quad \text{are independent.} \tag{3.20}$$

We are interested in testing

$$H_0: \ \mu = \mu^* \tag{3.21}$$

against the alternative

$$H_A: \ \mu \ne \mu^*. \tag{3.22}$$

The equality in (3.21) is in the space $L^2 = L^2([0,1])$, i.e. $\mu = \mu^*$ means that $\int (\mu(t) - \mu^*(t))^2 \, dt = 0$, and the alternative means that $\int (\mu(t) - \mu^*(t))^2 \, dt > 0$.

Since the statistical inference is about the mean functions of the observations, our procedures are based on the sample mean curves

$$\bar{X}_N(t) = \frac{1}{N} \sum_{1 \le i \le N} X_i(t) \quad \text{and} \quad \bar{X}_M^* = \frac{1}{M} \sum_{1 \le i \le M} X_j(t).$$

The sample means $\bar{X}_N$ and $\bar{X}_M^*$ are unbiased estimators of μ and μ^*, respectively, so H_0 will be rejected if

$$U_{N,M} = \frac{NM}{N+M} \int \left(\bar{X}_N(t) - \bar{X}_M^*(t)\right)^2 \, dt$$

is large.

Before introducing the test procedures, we state two results which describe the asymptotic behavior of the statistic $U_{N,M}$ under H_0 and H_A. They motivate and explain the development that follows.

Theorem 3.3. *Suppose H_0, the assumptions of Theorem 3.1 (and analogous assumptions for the ε_j^*) and (3.20) hold. If*

$$\frac{N}{N+M} \to \theta, \quad \text{for some } \ 0 \le \theta \le 1, \quad \text{as } \min(M,N) \to \infty, \tag{3.23}$$

then

$$U_{N,M} \xrightarrow{d} \int_0^1 \Gamma^2(t) \, dt,$$

where $\{\Gamma(t),\ 0 \le t \le 1\}$ is a mean zero Gaussian process with covariances

$$E[\Gamma(t)\Gamma(s)] = d(t,s) := (1-\theta)c(t,s) + \theta c^*(t,s).$$

Theorem 3.4. *If H_A, and the remaining assumptions of Theorem 3.3 hold, then*

$$\frac{N+M}{NM} U_{N,M} \xrightarrow{P} \int_0^1 \left(\mu(t) - \mu^*(t)\right)^2 \, dt.$$

In particular, if $0 < \theta < 1$, then $U_{N,M} \xrightarrow{P} \infty$.

The kernel $d(t,s)$ in Theorem 3.3 defines a covariance operator D. The eigenvalues of D are nonnegative, and are denoted by $\lambda_1 \geq \lambda_2 \geq \ldots$. By the Karhunen-Loéve expansion, we have

$$\int_0^1 \Gamma^2(t)\, dt = \sum_{i=1}^{\infty} \lambda_i N_i^2, \tag{3.24}$$

where $\{N_i,\ 1 \leq i < \infty\}$ are independent standard normal random variables.

Since the eigenvalues λ_i are unknown, the right-hand side of (3.24) cannot be used directly to simulate the distribution of $\int \Gamma^2(t)\, dt$. We will now explain how to estimate the λ_i's.

Suppose $\hat{D}_{N,M}$ is an L^2-consistent estimator of D , i.e.

$$\iint \left(\hat{d}_{N,M}(t,s) - d(t,s)\right)^2 dt\, ds \xrightarrow{P} 0, \quad \text{as } \min(M,N) \to \infty. \tag{3.25}$$

We discuss the construction of estimators $\hat{D}_{N,M}$ satisfying (3.25) below. For the estimators we propose, relation (3.25) holds regardless whether H_0 or H_A holds, they do not depend on μ or μ^* either. We will also see that the critical relations (3.45) hold under H_A as well as under H_0. The distribution of $\int \Gamma^2(t)\, dt$ can thus be estimated also under the alternative.

Let

$$\hat{\lambda}_1 = \hat{\lambda}_1(N,M) \geq \hat{\lambda}_2 = \hat{\lambda}_2(N,M) \geq \ldots$$

denote the eigenvalues of $\hat{D}_{N,M}$, i.e.

$$\int \hat{d}_{N,M}(t,s)\hat{\varphi}_i(s)\, ds = \hat{\lambda}_i \hat{\varphi}_i(t), \tag{3.26}$$

where the $\hat{\varphi}_i(t) = \hat{\varphi}_i(t; N, M)$ are the corresponding eigenfunctions satisfying $\int \hat{\varphi}_i^2(t)\, dt = 1$. Choosing p so large that $\sum_{i=1}^{p} \hat{\lambda}_i$ is a large percentage of $\sum_{i=1}^{N+M} \hat{\lambda}_i$, we can approximate the distribution of $\int \Gamma^2(t)\, dt$ by that of $\sum_{i=1}^{p} \hat{\lambda}_i N_i^2$.

The statistical inference is based on the difference $\bar{X}_N - \bar{X}_M^*$. Observe that

$$\frac{MN}{M+N} E\left[\left(\bar{X}_N(t) - \bar{X}_M^*(t)\right)\left(\bar{X}_N(s) - \bar{X}_M^*(s)\right)\right] \to d(t,s), \quad \text{as } \min(M,N) \to \infty,$$

that is, d is the asymptotic covariance kernel of the difference $\bar{X}_N - \bar{X}_M^*$. We therefore use projections onto the eigenfunctions $\varphi_1, \varphi_2, \ldots, \varphi_p$ associated with the p largest eigenvalues of D. This is analogous to projecting onto the functional principal components in one sample problems, as these form an L^2-optimal orthonormal basis. Without any loss of generality, we assume that the $\varphi_1, \varphi_2, \ldots, \varphi_p$ form an orthonormal system (the φ_i are orthogonal under (3.31), so only a normalization to unit norm is required). We define the projections

$$a_i = \left\langle \bar{X}_N - \bar{X}^*_M, \varphi_i \right\rangle, \quad 1 \le i \le p, \tag{3.27}$$

and the vectors

$$\mathbf{a} = [a_1, a_2, \ldots, a_p]^T.$$

We show in the proof of Theorem 3.5 that

$$\left(\frac{MN}{M+N} \right)^{1/2} \mathbf{a} \xrightarrow{d} \mathbf{N}_p(\mathbf{0}, \mathbf{Q}), \tag{3.28}$$

where $\mathbf{N}_p(\mathbf{0}, \mathbf{Q})$ stands for the p-variate normal random vector with mean zero and the covariance matrix $\mathbf{Q} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$. Since the operator D is unknown, we cannot compute the φ_i. However, any estimator for D satisfying (3.25) can be used to find estimates for the φ_i. Let $\hat{\varphi}_i$ be the empirical eigenfunctions defined by (3.26), and set

$$\hat{a}_i = \left\langle \bar{X}_N - \bar{X}^*_M, \hat{\varphi}_i \right\rangle, \quad 1 \le i \le p.$$

The limit relation (3.28) suggests the following statistics:

$$U^{(1)}_{N,M} = \frac{MN}{M+N} \sum_{i=1}^{p} \hat{a}_i^2 \tag{3.29}$$

and

$$U^{(2)}_{N,M} = \frac{MN}{M+N} \sum_{i=1}^{p} \frac{\hat{a}_i^2}{\hat{\lambda}_i}. \tag{3.30}$$

The following theorem establishes the limits of $U^{(1)}_{N,M}$ and $U^{(1)}_{N,M}$ under H_0.

Theorem 3.5. *Suppose H_0, the remaining assumptions of Theorem 3.3, (3.25) and*

$$\lambda_1 > \lambda_2 > \ldots > \lambda_p > \lambda_{p+1} \tag{3.31}$$

hold. Then

$$U^{(1)}_{N,M} \xrightarrow{d} \sum_{i=1}^{p} \lambda_i N_i^2, \tag{3.32}$$

where $N_1, N_2, \ldots, N_p$ are independent standard normal random variables, and

$$U^{(2)}_{N,M} \xrightarrow{d} \chi^2(p), \tag{3.33}$$

where $\chi^2(p)$ stands for a chi-square random variable with p degrees of freedom.

We note that $U_{N,M}^{(1)}$ is essentially the first p terms in the Karhunen-Loéve expansion of the integral in the definition of $U_{N,M}$. Thus, the limit in (3.32) is exactly the random variable we used to approximate the distribution of $U_{N,M}$. The limit in (3.33) is distribution free.

The consistency of the tests based on $U_{N,M}^{(1)}$ and $U_{N,M}^{(2)}$ follows from the following result.

Theorem 3.6. *Suppose H_A, the remaining assumptions of Theorem 3.3, (3.25) and (3.31) hold. Then*

$$\frac{N+M}{NM} U_{N,M}^{(1)} \xrightarrow{P} \sum_{i=1}^{p} \langle \mu - \mu^*, \varphi_i \rangle^2$$

and

$$\frac{N+M}{NM} U_{N,M}^{(2)} \xrightarrow{P} \sum_{i=1}^{p} \frac{\langle \mu - \mu^*, \varphi_i \rangle^2}{\lambda_i}.$$

In particular, if $0 < \theta < 1$ in (3.23), then $U_{N,M}^{(1)} \xrightarrow{P} \infty$ and $U_{N,M}^{(2)} \xrightarrow{P} \infty$, provided $\langle \mu - \mu^*, \varphi_i \rangle \neq 0$ for at least one $1 \leq i \leq p$.

We see that the condition for the consistency is that $\mu - \mu^*$ is not orthogonal to the linear subspace of L^2 spanned by the eigenfunctions $\varphi_i,\ 1 \leq i \leq p$.

The implementation of the tests based on Theorems 3.5 and 3.6 depends on the existence of an estimator of the kernel $d(t,s)$ which satisfies (3.25). The remainder of this section is dedicated to this issue.

The estimation of D is very simple if the ε_i are iid, and the ε_j^* are iid. In this case, setting,

$$\hat{\theta} = \frac{N}{N+M},$$

we can use

$$\hat{d}_{N,M}(t,s) = (1-\hat{\theta})\hat{c}_N(t,s) + \hat{\theta}\hat{c}_M^*(t,s), \tag{3.34}$$

where

$$\hat{c}_N(t,s) = \frac{1}{N} \sum_{i=1}^{N} \left(X_i(t) - \bar{X}_N(t) \right) \left(X_i(s) - \bar{X}_N(s) \right);$$

$$\hat{c}_M^*(t,s) = \frac{1}{M} \sum_{j=1}^{M} \left(X_j^*(t) - \bar{X}_M^*(t) \right) \left(X_j^*(s) - \bar{X}_N^*(s) \right).$$

By condition (3.5), we can use the weak law of large numbers in a Hilbert space to establish (3.25). The estimation of D is much more difficult if only (3.2) is assumed, and its asymptotic justification relies on Theorem 3.2. Recall the definition of the estimator $\hat{c}_N(t,s)$ given in (3.15), and define the estimator $\hat{c}_M^*(t,s)$ fully analogously. Our estimator for $d(t,s)$

is then (3.34) with $\hat{c}_N(t,s)$ and $\hat{c}^*_M(t,s)$ so defined. The following result then follows directly from Theorem 3.2.

Theorem 3.7. *Suppose the functional time series $\{X_i\}$ satisfies the assumptions of Theorem 3.2, and the series $\{X^*_j\}$ satisfies the same assumptions stated in terms of the ε^*_j. If (3.20) holds, then (3.25) holds.*

We emphasize that under the conditions of Theorem 3.7 relation (3.25) holds both under H_0 and H_A.

We now focus on the numerical issues related to the computation of the $\hat{a}_i$ and the $\hat{\zeta}_i$ appearing in the definitions of statistics $U^{(1)}_{N,M}$ and $U^{(2)}_{N,M}$. The $\hat{a}_i$ and the $\hat{\zeta}_i$ require the computation of the eigenfunctions and the eigenvalues of the operator $\hat{D}$. Except in the case of independent observations in each of the two samples, these quantities cannot be computed using existing software because $\hat{D}$ is not an empirical covariance operator of a functional iid sample. We recommend the following algorithm which we used to implement the tests. Let $\{e_\ell, \ell \geq 1\}$ be an orthonormal basis. The results reported in Section 3.4 are based on an implementation which uses the Fourier basis. In order to find approximate solutions to

$$\int \hat{d}_{N,M}(t,s)\phi(s)\,ds = \lambda\phi(t), \tag{3.35}$$

we use approximate expansions for $\phi(s)$ and $\hat{d}_{N,M}(t,s)$:

$$\phi(s) \approx \sum_{k=1}^{49} \phi_k e_k(t),$$

$$\hat{d}_{N,M}(t,s) \approx \sum_{k=1}^{49}\sum_{\ell=1}^{49} d_{k,\ell} e_k(t) e_\ell(s).$$

The coefficients ϕ_k and $d_{k,\ell}$ are given by

$$\phi_k = \int \phi(t) e_k(t)\,dt$$

and

$$d_{k,\ell} = \int\int \hat{d}_{N,M}(t,s) e_k(t) e_\ell(s)\,dt\,ds. \tag{3.36}$$

We replace $\phi(s)$ and $\hat{d}_{N,M}(t,s)$ in the left side of (3.35) with the above expansions to obtain

$$\sum_{k=1}^{49}\sum_{\ell=1}^{49} d_{k,\ell} e_k(t) \approx \lambda\phi(t). \tag{3.37}$$

Multiplying both sides of (3.37) by e_j and integrating yields

$$\sum_{\ell=1}^{49} d_{j,\ell}\phi_\ell \approx \lambda\phi_j, \quad 1 \le j \le 49.$$

In matrix form this is

$$\mathbf{D}\boldsymbol{\phi} = \lambda\boldsymbol{\phi},$$

where $\mathbf{D} = [d_{j,\ell},\ 1 \le j,\ell \le 49]$ and $\boldsymbol{\phi} = [\phi_1, \phi_2, \ldots, \phi_{49}]^T$. Thus we have reduced the problem of finding solutions of (3.35) to finding eigenvalues and eigenvectors of the matrix $\mathbf{D}$. Let $[\phi_{m,1}, \phi_{m,2}, \ldots, \phi_{m,49}]^T$ be the eigenvector corresponding to the m^{th} largest eigenvalue of $\mathbf{D}$. Then the eigenfunction associated with the m^{th} largest eigenvalue of $\hat{d}_{N,M}(t,s)$ is approximately $\sum_{\ell}^{49} \phi_{m,\ell} e_\ell(t)$. Using this notation, we obtain

$$\begin{aligned} \hat{a}_i &= \langle \bar{X}_N - \bar{X}_M^*, \hat{\varphi}_i \rangle \\ &\approx \sum_{\ell=1}^{49} \phi_{m,\ell} \left(N^{-1} \sum_{i=1}^{N} \langle X_i, e_\ell \rangle - M^{-1} \sum_{j=1}^{M} \langle X_j^*, e_\ell \rangle \right). \end{aligned} \tag{3.38}$$

We then obtain $U_{N,M}^{(1)}$ and $U_{N,M}^{(2)}$ as per (3.29) and (3.30).

To complete the description of the test procedures, we must specify how the value of p in the definitions of $U_{N,M}^{(1)}$ and $U_{N,M}^{(2)}$ is selected. This issue has been extensively studied in one sample problems, and several approaches have been put forward, including cross validation and penalty criteria. In our experience (for smooth densely recorded curves), the simple cumulative variance method advocated by Ramsay and Silverman (2005) has been satisfactory. We therefore recommend to use p such that the first p empirical functional principal components in each sample explain about 85% of the variance. As we will see in the data examples studied in Section 3.4, it is typically useful to look at the P-values for a range of p's.

3.4 A simulation study and data examples

We begin by presenting the results of a simulation study intended to evaluate the empirical size and power of the testing procedures introduced in Section 3.3. We then illustrate their properties on two data examples.

3.4.1 A simulation study

In this section, we compare the performance of the tests based on statistics $U_{N,M}^{(1)}$ and $U_{N,M}^{(2)}$ using simulated Gaussian functional data. We consider all combinations of sample sizes $N, M = 50, 100, 100$, and each pair of data generated processes was replicated three thousand times. To investigate the empirical size, without loss of generality, we set $\mu(t) = \mu^*(t) = 0$. Under the alternative, we set $\mu(t) = 0$ and $\mu^*(t) = at(1-t)$. The power is then a function of the parameter a. We considered two settings for the errors:

1. Both the $\varepsilon_i(t)$ and the $\varepsilon_j^*(t)$ are iid Brownian bridges.

2. Both the $\varepsilon_i(t)$ and the $\varepsilon_j^*(t)$ are functional AR(1) (FAR(1)) processes with the kernel,

$$\psi(t,s) = \frac{e^{-(t^2+s^2)/2}}{4\int_0^1 e^{-x^2}\,dx}.$$

   That is, the error terms, $\varepsilon_i(t)$, follow the model

$$\varepsilon_i(t) = \int_0^1 \psi(t,s)\varepsilon_{i-1}(s)\,ds + B_i(t),$$

   where $B_i(t)$ are iid Brownian bridges.

We calculated the test statistics $U_{N,M}^{(1)}$ and $U_{N,M}^{(2)}$ as explained in Section 3.3. These statistics depend on the choice of the weight functions K and K^*, and the bandwidth functions h and h^*. A great deal of attention has been devoted over several decades to the optimal selection of these functions for scalar and vector-valued time series, and we cannot address this issue within the space of this chapter. We follow the recommendation of Politis and Romano (1996) and use, for both samples, the flat top kernel

$$K(t) = \begin{cases} 1 & 0 \le |t| < 0.1, \\ 1.1 - |t| & 0.1 \le |t| < 1.1, \\ 0 & 1.1 \le |t| \end{cases}$$

with $h = N^{1/3}$ and $h^* = M^{1/3}$. We emphasize that this full estimation procedure was used for all data generating processes, including those with independent errors.

The results of the simulation study can be summarized as follows. The empirical size of the tests is larger in the case of FAR(1) errors. When a increases to 0.2 or larger, the empirical power of the test is smaller in the case of FAR(1) errors. Thus increasing the dependence in the error terms increases the size and decreases the power of the test. In both cases the tests have a slightly larger-than-nominal size and very good power. These observations are illustrated in Tables 3.1 and 3.2. Based on the whole simulation study, we can conclude that the performance of both tests is better if the sample sizes N and M are about equal. For example, for $N = M = 100$, the empirical sizes are closer to the nominal sizes then in the case $N = 100, M = 200$ shown in Tables 3.1 and 3.2. The power is very high even for small sample sizes. This is illustrated in Figure 3.2 which shows the samples with $N = M = 50$ and with slightly different means ($a = 0.8$). Visual inspection does not readily lead to the conclusion that the samples in the left and right panels of Figure 3.2 have different means, yet our tests can detect it with a very high probability. None of the two test statistics clearly dominates the other for the simulated Gaussian data, but a difference in behavior can be see when the tests are applied to real data sets, to which we now turn.

3.4.2 Mediterranean fruit flies

In our first example, it can be assumed that the curves in each sample are independent, as they were obtained from a randomized experiment. The data set used in this example was kindly made available to us by Hans-Georg Müller. It was extensively studied in biological and statistical literature, see Müller and Stadtmüller (2005) and references therein. We consider 534 egg-laying curves (count of eggs per unit time interval) of medflies who lived at least 30 days. Each function is defined over an interval $[0, 30]$, and its value on day $t \leq 30$ is the count of eggs laid by fly i on that day. The 534 flies are classified into long-lived, i.e., those who lived longer than 44 days, and short-lived, i.e., those who died before the end of the 44th day after birth. In the data set, there are 256 short-lived, and 278 long-lived flies. This classification naturally defines two samples: *Sample 1:* the egg-laying curves $\{X_i(t),\ 0 < t \leq 30,\ i = 1, 2, \ldots, 256\}$ of the short-lived flies. *Sample 2:* the egg-laying curves $\{X_j^*(t),\ 0 < t \leq 30,\ j = 1, 2, \ldots, 278\}$ of the long-lived flies. The egg-laying curves are very irregular; Figure 3.3 and Figure 3.4 show 10 smoothed curves of short- and long-lived flies. The tests are applied to such smooth trajectories.

Table 3.3 shows the P-values as a function of p. For both samples, $p = 2$ explains slightly over 85% of the variance, so this is the value we would recommend using. Both tests reject the equality of the mean functions, even though the sample means, shown in Figure 3.5, are not far apart. The P-values for the statistic $U^{(1)}$ are much more stable, equal to about 1%, no matter the value of p. The behavior of the test based on $U^{(2)}$ is more erratic. This indicates that while the test based on $U^{(2)}$ is easier to apply because it uses standard chi-square critical values, the test based on $U^{(1)}$ may be more reliable.

3.4.3 Eurodollar futures contracts

Our next example uses financial data kindly made available by Vladislav Kargin. This data is used as an example of modeling with the functional AR(1) process in Kargin and Onatski (2008). The curves, one curve per day, are constructed from the prices of

Table 3.1. Power of test (in %) using $U^{(1)}_{100,200}$ and $U^{(2)}_{100,200}$ with iid Brownian-bridge errors.

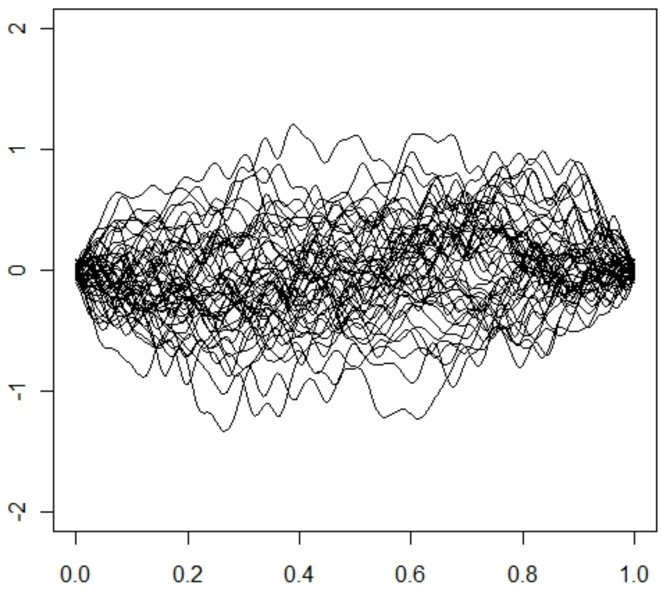
	$\alpha = .01$		$\alpha = .05$		$\alpha = .10$	
a	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$
0.0	1.5	1.5	6.3	6.2	11.4	11.6
0.1	2.5	3	7.4	8.0	13.2	13.6
0.2	6.0	4.4	16.7	13.0	24.8	20.2
0.3	14.2	9.2	30.4	23.2	41.3	33.0
0.4	26.4	17.0	48.5	36.1	60.8	48.0
0.5	44.0	31.4	64.7	53.5	75.2	64.3
0.6	59.4	45.9	80.3	68.0	87.9	78.4
0.7	78.0	64.7	91.8	82.4	96.0	89.0
0.8	88.0	78.0	95.9	90.9	98.0	94.6
0.9	94.2	88.1	98.7	95.8	99.4	97.9
1.0	98.0	94.8	99.5	98.5	99.9	99.3
1.1	99.6	98.4	100.0	99.8	100.0	100.0
1.2	99.9	99.4	100.0	99.9	100.0	100.0
1.3	100.0	99.8	100.0	100.0	100.0	100.0

Table 3.2. Power of test (in %) using $U^{(1)}_{100,200}$ and $U^{(2)}_{100,200}$ with FAR(1) errors.

	$\alpha = .01$		$\alpha = .05$		$\alpha = .10$	
a	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$
0.0	1.8	1.9	6.6	7.2	12.2	13.5
0.1	2.4	2.2	7.9	7.7	13.5	14.5
0.2	5.1	3.3	13.6	11.6	21.6	18.7
0.3	9.8	6.3	23.6	17.6	34.5	26.8
0.4	19.4	12.3	35.9	26.5	46.7	36.3
0.5	26.8	19.5	47.9	38.6	60.4	49.7
0.6	42.1	29.6	62.2	51.8	73.1	62.5
0.7	56.4	42.8	75.4	63.8	83.2	74.0
0.8	68.6	53.8	85.7	74.6	91.5	83.1
0.9	80.8	67.6	92.7	85.9	96.4	91.9
1.0	87.4	78.7	95.9	90.8	98.1	94.5
1.1	93.7	86.8	97.9	95.8	99.1	97.6
1.2	97.6	93.7	99.5	98.1	99.8	99.2
1.3	98.5	96.4	99.7	98.9	99.9	99.6

Table 3.3. P-values (in percent) of the tests based on statistics $U^{(2)}_{N,M}$ and $U^{(1)}_{N,M}$ applied to medfly data.

p	1	2	3	4	5	6	7	8	9
$U^{(1)}$	1.0	1.0	1.0	1.1	1.1	1.0	1.0	1.1	1.1
$U^{(2)}$	1.0	2.2	3.0	5.7	10.3	15.3	3.2	2.7	5.0



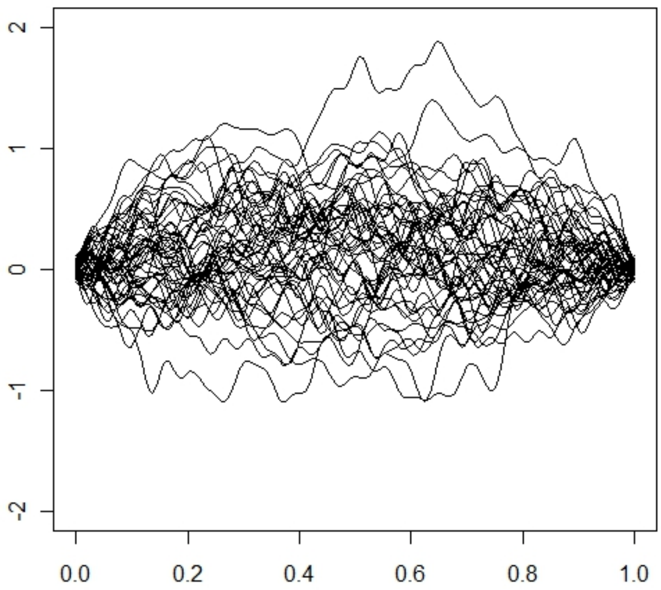


Figure 3.2. Fifty trajectories of the Brownian bridge (top) and 50 independent trajectories of the Brownian bridge plus $\mu^*(t) = 0.8t(1-t)$ (bottom). The tests can detect the different means with probability close to 90%.

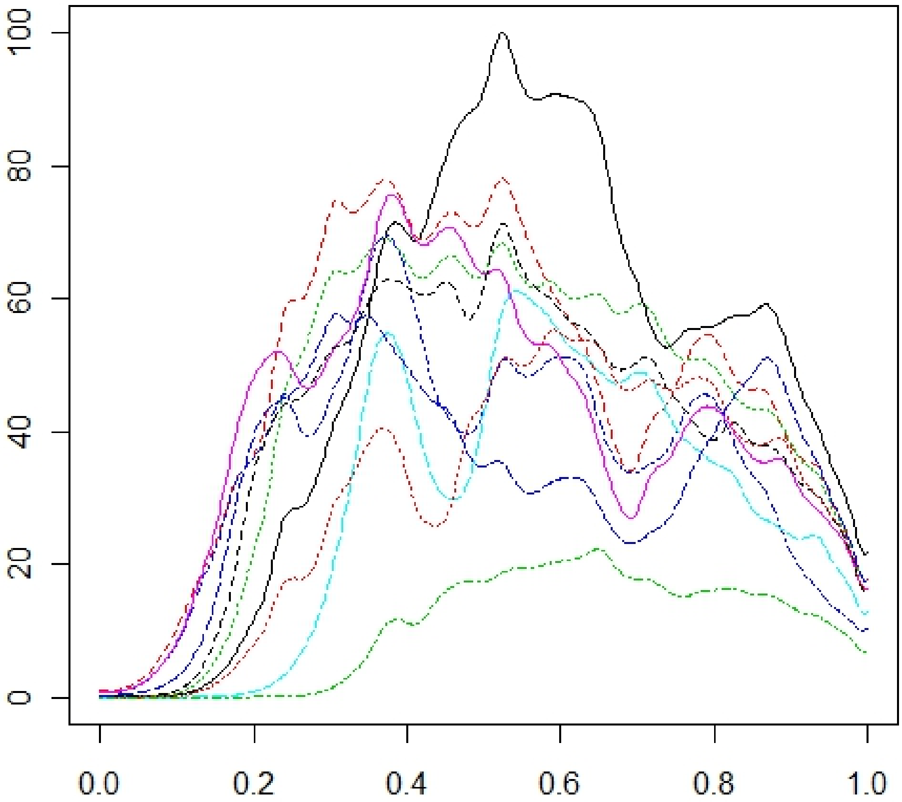


Figure 3.3. Ten randomly selected smoothed egg-laying curves of short-lived medflies.

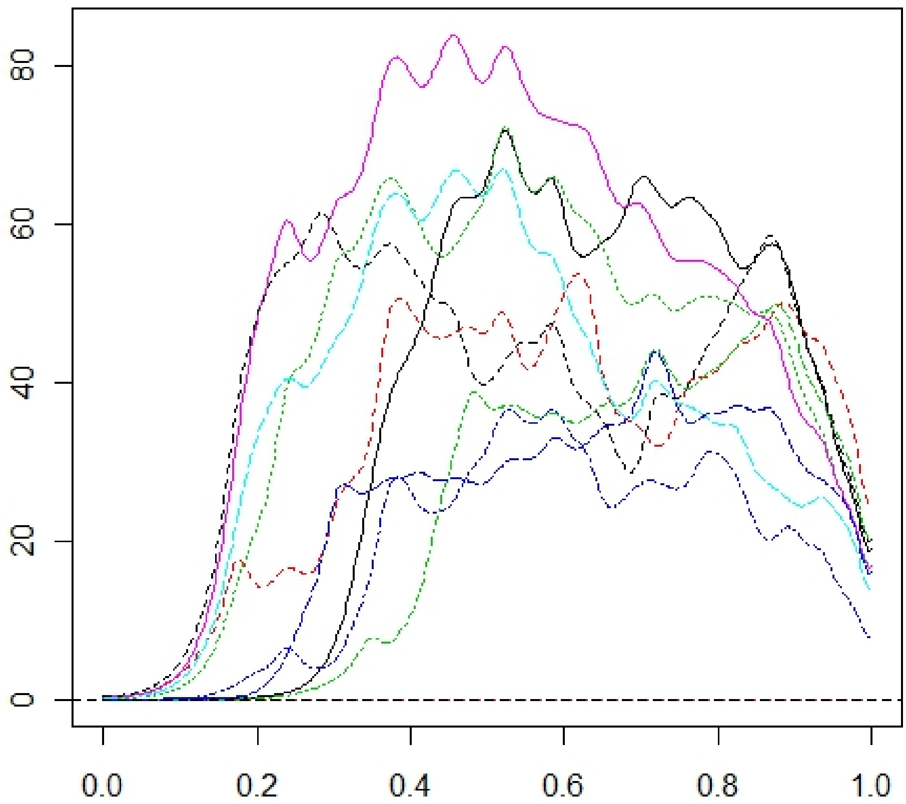


Figure 3.4. Ten randomly selected smoothed egg-laying curves of long-lived medflies.

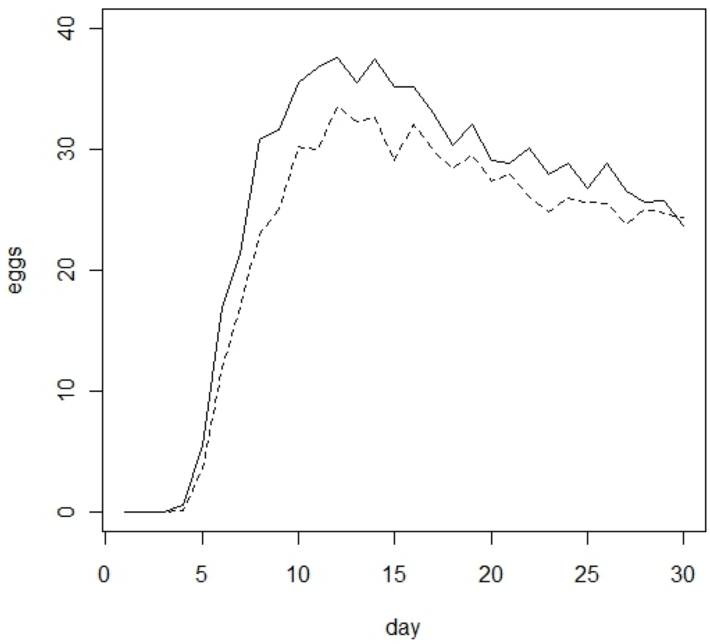


Figure 3.5. Estimated mean functions for the medfly data: short lived –solid line; long lived –dashed line.

Eurodollar futures contracts with decreasing expiration dates. The seller of a Eurodollar futures contract takes on an obligation to deliver a deposit of one million US dollars to a bank account outside the United States i months from today. The price the buyer is willing to pay for this contract depends on the prevailing interest rate. These contracts are traded on the Chicago Mercantile Exchange and provide a way to lock in an interest rate. Eurodollar futures are a liquid asset and are responsive to the Federal Reserve policy, inflation, and economic indicators.

The data we study consist of 114 points per day; point i corresponds to the price of a contract with closing date i months from today. We consider four samples, each consisting of 100 days of this data:

Sample 1: curves from September 7, 1999 to January 27, 2000.

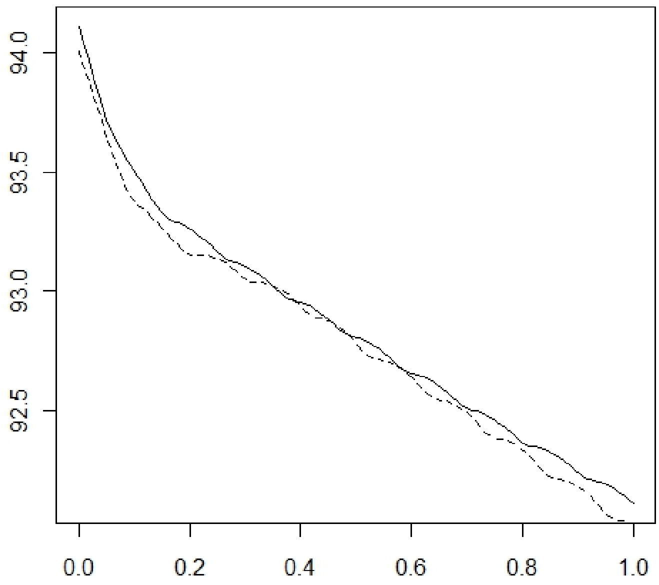
Sample 2: curves from January 24, 1997 to June 17, 1997.

Sample 3: curves from December 4, 1995 to April 24, 1996.

Sample 4: curves from March 6, 2001 to July 26, 2001.

Figure 3.6 shows the sample mean functions for the four samples. If a significance test does not reject H_0, we can conclude that the expectations of the future evolution of interest rates are the same for the two periods over which the samples were taken. A rejection means that these expectations are significantly different. As the analysis below reveals, we can conclude that expectations of future interest rates were different in Spring 1996 than in Summer 2001.

Table 3.4 shows the P-values as a function of p when the test is applied to samples 1 and 2, and also when it is applied to samples 3 and 4. In both samples 1 and 2, $p = 1$ explains more than 94% of the variance, in both samples 3 and 4, $p = 1$ explains more than 84% of the variance. Thus, following the recommendation of Section 3.3, we use the P-values obtained with $p = 1$. They lead to the acceptance of the null hypothesis of the equality of mean functions for periods corresponding to samples 1 and 2, and to its rejection for periods corresponding to samples 3 and 4 (notice that the 0.81 in the bottom panel of Table 3.4 is 0.81%). These conclusions agree with a visual evaluation of the sample mean functions in Figure 3.6. They also confirm the observation made in Section 3.4.1 that the tests have very good power, as the curves in the right panel of Figure 3.6 are not far apart. Both graphs in Figure 3.6 give us an idea what kind of differences in the sample mean functions are statistically significant, and which are not.



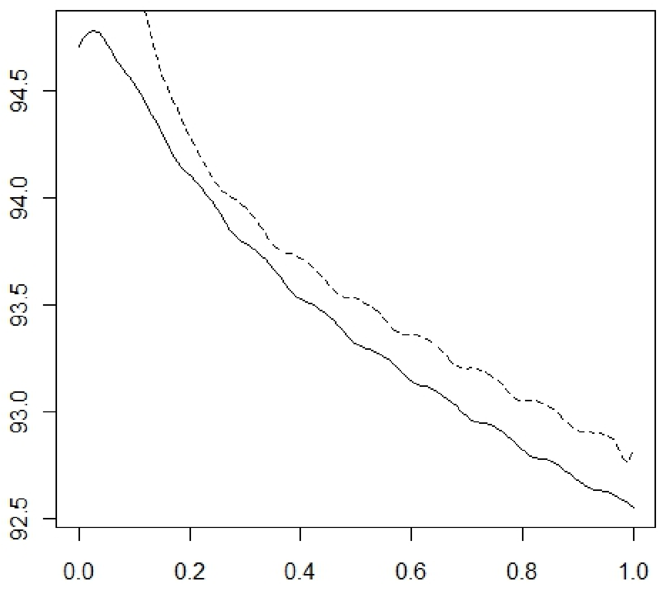


Figure 3.6. Sample means of the Eurodollar curves: Top: sample 1 solid; sample 2 dashed. Bottom: sample 3 solid; sample 4 dashed.

Table 3.4. P-values (in percent) of the statistics applied to Eurodollar data; samples 1 and 2 (top), samples 3 and 4 (bottom).

p	1	2	3	4	5
$U^{(1)}$	38.49	39.17	37.12	37.15	35.50
$U^{(2)}$	38.49	68.52	0.00	0.00	0.00

p	1	2	3	4	5
$U^{(1)}$	0.81	0.23	0.10	0.01	0.07
$U^{(2)}$	0.81	0.00	0.00	0.00	0.00

3.4.4 Conclusions

The simulations and data examples of this section show the tests we propose enjoy good finite sample properties. Tests of this type allow us to quantify statistical significance of conjectures made on the basis of exploratory analysis. For example, the sample mean curves in Figure 3.5 and the right panel of Figure 3.6 look a bit different, but a significance test allows us to state with confidence that they correspond to different population mean functions.

In many procedures of functional data analysis, both exploratory and inferential, the issue of choosing an optimal dimension reduction parameter, like the p in our setting, is delicate. Therefore, procedures less sensitive to such a choice are preferable. From this angle, the Monte Carlo test based $U^{(1)}$ is preferable, as an inspection of Tables 3.3 and 3.4 reveals. The test based on $U^{(2)}$ is however easier to apply, and in our examples and simulations leads to the same conclusions if p is chosen according to the cumulative variance rule.

The data examples of this section also show that the optimal value of p is typically a small single digit number, 1 or 2 in our examples. Therefore, developing asymptotics as p tends to infinity is not necessary, and may, in fact, be misleading because for larger values of p the tests may yield counterintuitive results.

3.5 Proofs of the results of Section 3.2

PROOF OF THEOREM 3.1. The proof is done in two steps. First we show that $N^{-1/2}\sum_{i=1}^N \varepsilon_i(t)$ is close to $N^{-1/2}\sum_{i=1}^N \varepsilon_{i,m}(t)$, if m is sufficiently large. Then we establish (3.6) for m-dependent functions for any $m \geq 1$.

As the first step, we show that

$$\limsup_{m\to\infty}\limsup_{N\to\infty} E\int\left[N^{-1/2}\sum_{i=1}^N\left(\varepsilon_i(t)-\varepsilon_{i,m}(t)\right)\right]^2 dt = 0, \tag{3.39}$$

where the variables $\varepsilon_{i,m}$ are defined in (3.3). By stationarity,

$$\begin{aligned}
&E\left[\sum_{1\leq i\leq N}\left(\varepsilon_i(t)-\varepsilon_{i,m}(t)\right)\right]^2\\
&=\sum_{1\leq i\leq N}\sum_{1\leq j\leq N} E\left(\varepsilon_i(t)-\varepsilon_{i,m}(t)\right)\left(\varepsilon_j(t)-\varepsilon_{j,m}(t)\right)\\
&=NE\left(\varepsilon_0(t)-\varepsilon_{0,m}(t)\right)^2+2\sum_{1\leq i<j\leq N} E\left(\varepsilon_i(t)-\varepsilon_{i,m}(t)\right)\left(\varepsilon_j(t)-\varepsilon_{j,m}(t)\right).
\end{aligned}$$

In the proof, we will repeatedly use independence relations which follow from representations (3.1) and (3.3). First observe that if $j > i$, then $(\varepsilon_i, \varepsilon_{i,m})$ is independent of $\varepsilon_{j,j-i}$ because

$$\varepsilon_{j,j-i} = f(\delta_j, \ldots, \delta_{i+1}, \delta_{j,i}^{(j-i)}, \delta_{j,i-1}^{(j-i)}, \ldots).$$

Consequently, $E\left(\varepsilon_i(t) - \varepsilon_{i,m}(t)\right)\varepsilon_{j,j-i}(t) = 0$, and so

$$\sum_{1\le i<j\le N} E\left(\varepsilon_i(t) - \varepsilon_{i,m}(t)\right)\varepsilon_j(t) = \sum_{1\le i<j\le N} E\left(\varepsilon_i(t) - \varepsilon_{i,m}(t)\right)\left(\varepsilon_j(t) - \varepsilon_{j,j-i}(t)\right).$$

Using the Cauchy-Schwarz inequality and (3.2), we conclude

$$\begin{aligned}
&\left|\int \sum_{1\le i<j\le N} E\left(\varepsilon_i(t) - \varepsilon_{i,m}(t)\right)\left(\varepsilon_j(t) - \varepsilon_{j,j-i}(t)\right)\, dt\right| \\
&\qquad \le \sum_{1\le i<j\le N} \int \left[E\left(\varepsilon_i(t) - \varepsilon_{i,m}(t)\right)^2\right]^{1/2} \left[E\left(\varepsilon_j(t) - \varepsilon_{j,j-i}(t)\right)^2\right]^{1/2} dt \\
&\qquad \le \sum_{1\le i<j\le N} \left[\int E\left(\varepsilon_i(t) - \varepsilon_{i,m}(t)\right)^2 dt\right]^{1/2} \left[\int E\left(\varepsilon_j(t) - \varepsilon_{j,j-i}(t)\right)^2 dt\right]^{1/2} \\
&\qquad = \sum_{1\le i<j\le N} \left[\int E\left(\varepsilon_0(t) - \varepsilon_{0,m}(t)\right)^2 dt\right]^{1/2} \left[\int E\left(\varepsilon_0(t) - \varepsilon_{0,j-i}(t)\right)^2 dt\right]^{1/2} \\
&\qquad \le N \left[\int E\left(\varepsilon_{0,m}(t) - \varepsilon_0(t)\right)^2 dt\right]^{1/2} \sum_{k\ge 1} \left[\int \left(\varepsilon_0(t) - \varepsilon_{0,k}(t)\right)^2 dt\right]^{1/2}.
\end{aligned}$$

Hence

$$\limsup_{m\to\infty} \limsup_{N\to\infty} \frac{1}{N} \left|\int \sum_{1\le i<j\le N} E\left[\left(\varepsilon_{i,m}(t) - \varepsilon_i(t)\right)\varepsilon_j(t)\right] dt\right| = 0.$$

Similar arguments give

$$\limsup_{m\to\infty} \limsup_{N\to\infty} \frac{1}{N} \left|\int \sum_{1\le i<j\le N} E\left[\left(\varepsilon_{i,m}(t) - \varepsilon_i(t)\right)\varepsilon_{j,m}(t)\right] dt\right| = 0.$$

Completing the verification of (3.39).

The next the step is to show that $N^{-1/2}\sum_{1\le i\le N} \varepsilon_{i,m}$ converges to a Gaussian process Z_m with covariances defined analogously to (3.7). Recall that for every integer $m \ge 1$, $\{\varepsilon_{i,m}\}$ is an m-dependent sequence of functions. To lighten the notation, in the remainder of the proof, we fix m and denote sequence $\{\varepsilon_{i,m}\}$ by $\{\varepsilon_i\}$, so $\{\varepsilon_i\}$ is now m-dependent.

Let $K > 1$ be an integer and ψ_i be an orthonormal basis determined by the eigenfunctions of $E\varepsilon(t)\varepsilon(s)$. The corresponding eigenvalues are denoted by ν_i. Then, by the Karhunen-Loéve expansion, we have

$$\varepsilon_i(t) = \sum_{\ell \geq 1} \langle \varepsilon_i, \psi_\ell \rangle \, \psi_\ell(t).$$

Next we define

$$\varepsilon_i^{(K)}(t) = \sum_{1 \leq \ell \leq K} \langle \varepsilon_i, \psi_\ell \rangle \, \psi_\ell(t).$$

By the triangle inequality we have that

$$\begin{aligned}
&\left\{ E \int \left[\sum_{1 \leq i \leq N} \left((\varepsilon_i(t) - \varepsilon_i^{(K)}(t) \right) \right]^2 dt \right\}^{1/2} \\
&\quad \leq \left\{ E \int \left[\sum_{i \in V(0)} \left((\varepsilon_i(t) - \varepsilon_i^{(K)}(t) \right) \right]^2 dt \right\}^{1/2} + \ldots \\
&\qquad + \left\{ E \int \left[\sum_{i \in V(m-1)} \left((\varepsilon_i(t) - \varepsilon_i^{(K)}(t) \right) \right]^2 dt \right\}^{1/2},
\end{aligned}$$

where $V(k) = \{i : \ 1 \leq i \leq N, i = k \pmod m\}, 0 \leq k \leq m-1$. Due to the m dependence of the sequence $\{\varepsilon_i\}$, $\sum_{i \in V(k)} (\varepsilon_i(t) - \varepsilon_i^{(K)}(t))$ is a sum of independent, identically distributed random variables, and thus we get that

$$E \int \left[\sum_{i \in V(m-1)} \left((\varepsilon_i(t) - \varepsilon_i^{(K)}(t) \right) \right]^2 dt \leq N \sum_{\ell \geq K} E \, \langle X_0, \psi_\ell \rangle^2 .$$

Utilizing

$$\lim_{K \to \infty} \sum_{\ell \geq K} E \, \langle X_0, \psi_\ell \rangle^2 = 0$$

we conclude that for any $x > 0$

$$\limsup_{K \to \infty} \limsup_{N \to \infty} P \left\{ \int \left[\frac{1}{N^{1/2}} \sum_{1 \leq i \leq N} \left((\varepsilon_i(t) - \varepsilon_i^{(K)}(t) \right) \right]^2 dt > x \right\} = 0.$$

The sum of the $\varepsilon_i^{(K)}$'s can be written as

$$\frac{1}{N^{1/2}} \sum_{1 \leq i \leq N} \varepsilon_i^{(K)}(t) = \sum_{1 \leq \ell \leq K} \psi_\ell(t) \frac{1}{N^{1/2}} \sum_{1 \leq i \leq N} \langle \varepsilon_i, \psi_\ell \rangle .$$

Next, we use the central limit theorem for stationary m-dependent sequences of random vectors (see Lehmann (1999) and the Cramér-Wold theorems in DasGupta (2008), pages 9 and 120)) and get that

$$\left\{ \frac{1}{N^{1/2}} \sum_{1 \le i \le N} \langle \varepsilon_i, \psi_\ell \rangle, 1 \le \ell \le K \right\}^T \xrightarrow{d} \mathbf{N}_K(\mathbf{0}, \boldsymbol{\Delta}_K),$$

where $\mathbf{N}_K(\mathbf{0}, \boldsymbol{\Delta}_K)$ is a K-dimensional normal random variable with zero mean and covariance matrix $\boldsymbol{\Delta}_K = \mathrm{diag}(\nu_1, \ldots, \nu_K)$. Thus we proved that for all $K > 1$

$$N^{-1/2} \sum_{1 \le i \le N} \varepsilon_i^{(K)}(t) \xrightarrow{d} \sum_{1 \le \ell \le K} \nu_\ell^{1/2} N_\ell \psi_\ell(t) \quad \text{in} \;\; L^2,$$

where $N_i, i \ge 1$ are independent standard normal random variables. It is easy to see that

$$\int \left(\sum_{K < \ell < \infty} \nu_\ell^{1/2} N_\ell \psi_\ell(t) \right)^2 dt = \sum_{K < \ell < \infty} \nu_\ell N_\ell^2 \xrightarrow{P} 0,$$

as $K \to \infty$. Thus we have the convergence of $N^{-1/2} \sum_{1 \le i \le N} \varepsilon_i$ for any m and therefore the proof of the Theorem is now complete.

Proof of Theorem 3.2. First we reduce (3.18) to (3.40). Then, we reduce (3.40) to (3.44).

Since

$$\hat{\gamma}_0(t,s) = \frac{1}{N} \sum_{i=1}^{N} (X_i(t) - \mu(t))(X_i(s) - \mu(s)) - \left(\bar{X}_N(t) - \mu(t)\right)\left(\bar{X}_N(s) - \mu(s)\right),$$

we obtain that

$$\begin{aligned}
&\iint \{\hat{\gamma}_0(t,s) - E[\varepsilon_0(t)\varepsilon_0(s)]\}^2 \, dt\, ds \\
&\quad \le 4 \iint \left\{ \frac{1}{N} \sum_{i=1}^{N} (X_i(t) - \mu(t))(X_i(s) - \mu(s)) - E[\varepsilon_0(t)\varepsilon_0(s)] \right\}^2 dt\, ds \\
&\qquad + 4 \left(\int \left(\bar{X}_N(t) - \mu(t)\right)^2 dt \right)^2 \\
&\quad = o_P(1),
\end{aligned}$$

using the ergodic theorem for random variables in a Hilbert space.

Next observe that

$$\hat{\gamma}_i(t,s) = \frac{1}{N}\sum_{j=i+1}^{N} \varepsilon_j(t)\varepsilon_{j-i}(s)$$
$$+\frac{N-i}{N}\bar{\varepsilon}_N(t)\bar{\varepsilon}_N(s) - \left(\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\right)\bar{\varepsilon}_N(s) - \bar{\varepsilon}_N(t)\left(\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_{j-i}(s)\right).$$

Therefore, setting,

$$\bar{\gamma}_i(t,s) = \frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\varepsilon_{j-i}(s).$$

we obtain

$$\begin{aligned}
\sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)\hat{\gamma}_i(t,s) &= \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)\bar{\gamma}_i(t,s)\\
&\quad - \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)\left\{\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\right\}\bar{\varepsilon}_N(s)\\
&\quad - \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)\left\{\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_{j-i}(s)\right\}\bar{\varepsilon}_N(t)\\
&\quad + \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)\frac{N-i}{N}\bar{\varepsilon}_N(t)\bar{\varepsilon}_N(s).
\end{aligned}$$

By stationarity we conclude that for any $1 \le i \le N$,

$$\begin{aligned}
E\int\left(\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\right)^2 dt &= \frac{1}{N^2}\sum_{j=i+1}^{N}\int E\varepsilon_0^2(t)\,dt + \frac{2}{N^2}\sum_{j=i+1}^{N}(N-j)\int E\varepsilon_0(t)\varepsilon_j(t)\,dt\\
&\le \frac{1}{N}\int E\varepsilon_0^2(t)\,dt + \frac{4}{N}\sum_{j=1}^{\infty}\left|\int E\varepsilon_0(t)\varepsilon_j(t)\,dt\right|.
\end{aligned}$$

Since ε_0 and $\varepsilon_{j,j}$ are independent, we get by (3.2)

$$\begin{aligned}
\sum_{j=1}^{\infty}\left|\int E\varepsilon_0(t)\varepsilon_j(t)\,dt\right| &= \sum_{j=1}^{\infty}\left|\int E\varepsilon_0(t)\left(\varepsilon_j(t)-\varepsilon_{j,j}(t)\right)\,dt\right|\\
&\le \sum_{j=1}^{\infty} E\left\{\left(\int\varepsilon_0^2(t)\,dt\right)^{1/2}\left(\int\left(\varepsilon_j(t)-\varepsilon_{j,j}(t)\right)^2\,dt\right)^{1/2}\right\}\\
&\le \left(E\int\varepsilon_0^2(t)\,dt\right)^{1/2}\sum_{j=1}^{\infty}\left(\int E\left(\varepsilon_j(t)-\varepsilon_{j,j}(t)\right)^2\,dt\right)^{1/2}\\
&< \infty.
\end{aligned}$$

Thus, we have

$$\max_{1\le i\le N} E\int\left(\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\right)^2 dt = O(1).$$

Consequently, using the triangle inequality,

$$\begin{aligned}
&E\left\{\iint\left(\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\left[\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\right]\bar{\varepsilon}_N(s)\right)^2 dt\,ds\right\}^{1/2}\\
&\le \sum_{i=1}^{N-1}\left|K\left(\frac{i}{h}\right)\right| E\left\{\left(\int\left[\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\right]^2 dt\right)^{1/2}\left(\int\bar{\varepsilon}_N^2(s)\,ds\right)^{1/2}\right\}\\
&\le \sum_{i=1}^{N-1}\left|K\left(\frac{i}{h}\right)\right|\left(E\int\left[\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\right]^2\right)^{1/2}\left(\int\bar{\varepsilon}_N^2(s)\,ds\right)^{1/2}\\
&= \frac{h}{N}O(1) = o(1),
\end{aligned}$$

on account of (3.16).

Hence to establish (3.18) , it is enough to prove that

$$\iint\left(\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\bar{\gamma}_i(t,s) - c_1(t,s)\right)^2 dt\,ds = o_P(1), \tag{3.40}$$

where

$$c_1(t,s) = \sum_{i\ge 1}E[\varepsilon_0(s)\varepsilon_i(t)].$$

Let $\{\varepsilon_{n,m},\ -\infty < n < \infty\}$ be the random variables defined in (3.3), where m is a fixed number. Let

$$\tilde{\gamma}_{i,m}(t,s) = \frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_{j,m}(t)\varepsilon_{j-i,m}(s).$$

We show that for every $m \ge 1$,

$$\iint\left(\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\tilde{\gamma}_{i,m}(t,s) - c_1^{(m)}(t,s)\right)^2 dt\,ds = o_P(1), \tag{3.41}$$

where

$$c_1^{(m)}(t,s) = \sum_{i=1}^{\infty}E[\varepsilon_{1,m}(s)\varepsilon_{i+1,m}(t)].$$

We also note that (3.3) and (3.2) imply

$$\lim_{m\to\infty}\iint\left(c_1^{(m)}(t,s) - c_1(t,s)\right)^2 dt\,ds = 0. \tag{3.42}$$

Since $\{\varepsilon_{n,m},\ -\infty < n < \infty\}$ is an m-dependent sequence,

$$c_1^{(m)}(t,s) = \sum_{i=1}^{m} E[\varepsilon_{1,m}(s)\varepsilon_{i+1,m}(t)].$$

Using (3.10), (3.11) and (3.16), we get

$$\max_{1\le i\le m}\left|K\left(\frac{i}{h}\right)-1\right| \to 0, \quad \text{as } N\to\infty.$$

By the ergodic theorem,

$$\iint \left(\tilde{\gamma}_{i,m}(t,s) - E[\varepsilon_{1,m}(s)\varepsilon_{i+1,m}(t)]\right)^2 dt\,ds = o_P(1),$$

for any fixed i. Hence (3.41) is proven, once we have shown that

$$\iint \left(\sum_{i=m+1}^{N-1} K\left(\frac{i}{h}\right)\tilde{\gamma}_{i,m}(t,s)\right)^2 dt\,ds = o_P(1). \tag{3.43}$$

It is easy to see that

$$\begin{aligned} &E\iint \left(\sum_{i=m+1}^{N-1} K\left(\frac{i}{h}\right)\tilde{\gamma}_{i,m}(t,s)\right)^2 dt\,ds \\ &= \iint \left(\frac{1}{N^2}\sum_{i=m+1}^{h}\sum_{\ell=m+1}^{h}\sum_{k=i+1}^{N-1}\sum_{n=\ell+1}^{N-1} K\left(\frac{i}{h}\right)K\left(\frac{\ell}{h}\right)E\left[\varepsilon_{k,m}\varepsilon_{k-i,m}\varepsilon_{n,m}\varepsilon_{n-\ell,m}\right]\right), \end{aligned}$$

provided $h \le N-1$. The sequence $\{\varepsilon_{n,m},\ -\infty < n < \infty\}$ is an m-dependent, and therefore $\varepsilon_{k,m}$ and $\varepsilon_{k-i,m}$ are independent, since $i \ge m+1$. Similarly, $\varepsilon_{n,m}$ and $\varepsilon_{n-\ell,m}$ are independent. Hence the number of terms when $E\left[\varepsilon_{k,m}\varepsilon_{k-i,m}\varepsilon_{n,m}\varepsilon_{n-\ell,m}\right]$ is different from zero is $O(Nh)$. Consequently,

$$E\iint \left(\sum_{i=m+1}^{N-1} K\left(\frac{i}{h}\right)\tilde{\gamma}_{i,m}(t,s)\right)^2 dt\,ds = O\left(\frac{h}{N}\right) = o(1).$$

This completes the verification of(3.43).

Next we show that for all $\epsilon > 0$,

$$\lim_{m\to\infty}\limsup_{N\to\infty} P\left\{\iint \left(\sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)\left[\bar{\gamma}_i(t,s) - \tilde{\gamma}_{i,m}(t,s)\right]\right)^2 dt\,ds > \epsilon\right\} = 0. \tag{3.44}$$

Using the definitions of the covariances $\bar{\gamma}_i(t,s)$ and $\tilde{\gamma}_{i,m}(t,s)$, we consider the decompositions

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}\left[\varepsilon_j(t)\varepsilon_{j-i}(s) - \varepsilon_{j,m}(t)\varepsilon_{j-i,m}(s)\right] \\ &\qquad = \frac{1}{N}\left\{\sum_{i=1}^{m} + \sum_{i=m+1}^{h}\right\} K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}\left[\varepsilon_j(t)\varepsilon_{j-i}(s) - \varepsilon_{j,m}(t)\varepsilon_{j-i,m}(s)\right] \end{aligned}$$

and

$$\varepsilon_j(t)\varepsilon_{j-i}(s) - \varepsilon_{j,m}(t)\varepsilon_{j-i,m}(s) = (\varepsilon_j(t) - \varepsilon_{j,m}(t))\,\varepsilon_{j-i}(s) + (\varepsilon_{j-i}(s) - \varepsilon_{j-i,m}(s))\,\varepsilon_{j,m}(t).$$

Clearly,

$$\begin{aligned}
&\left\{\iint \left(\frac{1}{N}\sum_{i=1}^{m} K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}(\varepsilon_j(t) - \varepsilon_{j,m}(t))\,\varepsilon_{j-i}(s)\right)^2 dt\,ds\right\}^{1/2} \\
&\quad \le \frac{1}{N}\sum_{i=1}^{m}\left|K\left(\frac{i}{h}\right)\right|\left\{\int (\varepsilon_j(t) - \varepsilon_{j,m}(t))^2\,dt\right\}^{1/2}\left\{\int \varepsilon_{j-i}^2(s)\,ds\right\}^{1/2},
\end{aligned}$$

so, by (3.17),

$$\begin{aligned}
&E\left\{\iint \left(\frac{1}{N}\sum_{i=1}^{m} K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}(\varepsilon_j(t) - \varepsilon_{j,m}(t))\,\varepsilon_{j-i}(s)\right)^2 dt\,ds\right\}^{1/2} \\
&\quad \le m\left\{E\int (\varepsilon_0(t) - \varepsilon_{0,m}(t))^2\,dtE\int \varepsilon_0^2(s)\,ds\right\}^{1/2} \\
&\quad \le Am\left\{E\int (\varepsilon_0(t) - \varepsilon_{0,m}(t))^2\,dt\right\}^{1/2} \to 0, \quad \text{as} \quad m \to \infty,
\end{aligned}$$

according to (3.17), where A is a constant.

Next we use the decomposition

$$\varepsilon_j(t)\varepsilon_{j-i}(s) = \varepsilon_{j,i}(t)\varepsilon_{j-i}(s) + (\varepsilon_j(t) - \varepsilon_{j,i}(t))\,\varepsilon_{j-i}(s)$$

to get

$$\begin{aligned}
&\left\{\iint \left(\frac{1}{N}\sum_{i=m+1}^{h} K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}(\varepsilon_j(t) - \varepsilon_{j,i}(t))\,\varepsilon_{j-i}(s)\right)^2 dt\,ds\right\}^{1/2} \\
&\quad \le \frac{1}{N}\sum_{i=m+1}^{\infty}\sum_{j=i+1}^{N}\left\{\int (\varepsilon_j(t) - \varepsilon_{j,i}(t))^2\,dt\right\}^{1/2}\left\{\int \varepsilon_{j-i}^2(s)\,ds\right\}^{1/2}.
\end{aligned}$$

Therefore, by (3.5) and (3.2), we have

$$\begin{aligned}
&E\left\{\iint \left(\frac{1}{N}\sum_{i=m+1}^{h} K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}(\varepsilon_j(t) - \varepsilon_{j,i}(t))\,\varepsilon_{j-i}(s)\right)^2 dt\,ds\right\}^{1/2} \\
&\le \frac{1}{N}\sum_{i=m+1}^{\infty}\sum_{j=i+1}^{N} E\left[\left\{\int (\varepsilon_j(t) - \varepsilon_{j,i}(t))^2\,dt\right\}^{1/2}\left\{\int \varepsilon_{j-i}^2(s)\,ds\right\}^{1/2}\right] \\
&\le \frac{1}{N}\sum_{i=m+1}^{\infty}\sum_{j=i+1}^{N}\left[\int (\varepsilon_j(t) - \varepsilon_{j,i}(t))^2\,dt\right]^{1/2}\left[\int \varepsilon_{j-i}^2(s)\,ds\right]^{1/2} \\
&\le A\sum_{i=m+1}^{\infty}\left[\int (\varepsilon_0(t) - \varepsilon_{0,i}(t))^2\,dt\right]^{1/2} \to 0, \quad \text{as} \quad m \to \infty.
\end{aligned}$$

We have shown so far that for any $\epsilon > 0$,

$$\lim_{m\to\infty} \limsup_{N\to\infty} P\left\{ \iint \left(\frac{1}{N} \sum_{i=m+1}^{h} K\left(\frac{i}{h}\right) \sum_{j=i+1}^{N} \varepsilon_j(t)\varepsilon_{j-i}(s) \right)^2 dt\, ds > \epsilon \right\} = 0.$$

Similar arguments give

$$\lim_{m\to\infty} \limsup_{N\to\infty} P\left\{ \iint \left(\frac{1}{N} \sum_{i=m+1}^{h} K\left(\frac{i}{h}\right) \sum_{j=i+1}^{N} \varepsilon_{j,m}(t)\varepsilon_{j-i,m}(s) \right)^2 dt\, ds > \epsilon \right\} = 0.$$

This completes the verification of (3.44), so (3.40) is proven.

3.6 Proofs of the results of Section 3.3

In the proofs, we will often use the relations

$$\max_{1\le i\le p} |\hat{\lambda}_i - \lambda_i| \xrightarrow{P} 0, \quad \text{and} \quad \max_{1\le i\le p} ||\hat{\varphi}_i - \hat{c}_i\varphi_i|| \xrightarrow{P} 0 \quad \text{as } \min(M,N) \to \infty, \tag{3.45}$$

where $\hat{c}_i = \operatorname{sign}(\langle \hat{\varphi}_i, \varphi_i \rangle)$, Analogous relations have been extensively used for the eigenvalues and the eigenfunctions of the empirical and population covariance operators, see Bosq (2000), Gabrys and Kokoszka (2007), Horváth et al. (2010), Panaretos et al. (2010), among many others, but they hold in much greater generality, see Chapter 2 of Horváth and Kokoszka (2011). Under (3.31), they hold for the eigenelements of the operators $\hat{D}$ and D defined in Section 3.3.

Proof of Theorem 3.3: By Theorem 3.1, assumptions (3.1)–(3.2) imply that, as $N \to \infty$ and $M \to \infty$

$$\left(N^{-1/2} \sum_{1\le i\le N} \varepsilon_i,\ M^{-1/2} \sum_{1\le j\le N} \varepsilon_j^* \right) \xrightarrow{d} \left(\Gamma^{(1)}, \Gamma^{(2)} \right),$$

where $\Gamma^{(1)}$ and $\Gamma^{(2)}$ are independent L^2-valued mean zero Gaussian processes with covariances

$$E\left[\Gamma^{(1)}(t)\Gamma^{(1)}(s) \right] = c(t,s), \quad E\left[\Gamma^{(2)}(t)\Gamma^{(2)}(s) \right] = c^*(t,s).$$

Hence

$$\frac{NM}{N+M} \int \left(N^{-1} \sum_{1\le i\le N} \varepsilon_i(t) - M^{-1} \sum_{1\le j\le M} \varepsilon_j^*(t) \right)^2 dt \xrightarrow{d} \int \Gamma(t)\, dt, \tag{3.46}$$

where $\Gamma(t) = (1-\theta)^{1/2}\Gamma^{(1)}(t) + \theta^{1/2}\Gamma^{(2)}(t)$. The conclusion of Theorem 3.3 now follows. $\square$

Proof of Theorem 3.4: By the ergodic theorem in a Hilbert space, $||\bar{X}_N - \mu|| = o_P(1)$ and $||\bar{X}_M - \mu^*|| = o_P(1)$, which imply the result. $\square$

Proof of Theorem 3.5: Under assumptions (3.5) and (3.2), we have, jointly,

$$\left\{ N^{-1/2} \sum_{1 \le i \le N} \langle \varepsilon_i, \varphi_k \rangle\,,\ 1 \le k \le p \right\} \xrightarrow{d} \mathbf{N}_p^{(1)}\left(\mathbf{0}, \mathbf{Q}^{(1)}\right)$$

and

$$\left\{ M^{-1/2} \sum_{1 \le j \le M} \left\langle \varepsilon_j^*, \varphi_k \right\rangle\,,\ 1 \le k \le p \right\} \xrightarrow{d} \mathbf{N}_p^{(2)}\left(\mathbf{0}, \mathbf{Q}^{(2)}\right),$$

where $\mathbf{N}_p^{(1)}\left(\mathbf{0}, \mathbf{Q}^{(1)}\right)$ and $\mathbf{N}_p^{(2)}\left(\mathbf{0}, \mathbf{Q}^{(2)}\right)$ are independent p-dimensional normal random vectors. Since

$$a_k = N^{-1} \sum_{1 \le i \le N} \langle \varepsilon_i, \varphi_k \rangle - M^{-1} \sum_{1 \le j \le M} \left\langle \varepsilon_j^*, \varphi_k \right\rangle. \quad 1 \le i \le p,$$

(3.28) holds with $\mathbf{Q} = (1-\theta)\mathbf{Q}^{(1)} + \theta\mathbf{Q}^{(2)}$.

Next we observe that the matrix $\mathbf{Q} = (Q(k,\ell),\ 1 \le k, \ell \le p)$ satisfies

$$Q(k,\ell) = \iint d(t,s)\varphi_k(t)\varphi_\ell(s)\,dt\,ds = \lambda_k \delta_{k\ell},$$

where δ_{ij} is Kronecker's delta, using the fact that the φ_i are orthonormal eigenfunctions and the λ_i are the corresponding eigenvalues.

Recall now relations (3.45), and observe that

$$\begin{aligned} \hat{a}_k = \hat{c}_k &\left(N^{-1} \sum_{1 \le i \le N} \langle \varepsilon_i, \varphi_k \rangle - M^{-1} \sum_{1 \le j \le M} \left\langle \varepsilon_j^*, \varphi_k \right\rangle \right) \\ &+ \left\langle N^{-1} \sum_{1 \le i \le N} \varepsilon_i - M^{-1} \sum_{1 \le j \le M} \varepsilon_j^*\ ,\ \hat{\varphi}_k - \hat{c}_k \varphi_k \right\rangle. \end{aligned}$$

By (3.45) and (3.46),

$$\begin{aligned} \left(\frac{NM}{N+M} \right)^{1/2} &\left| \left\langle \frac{1}{N} \sum_{1 \le i \le N} \varepsilon_i - \frac{1}{M} \sum_{1 \le j \le M} \varepsilon_j^*\ ,\ \hat{\varphi}_k - \hat{c}_k \varphi_k \right\rangle \right| \\ &\le U_{N,M}^{1/2} \left(\int (\hat{\varphi}_k - \hat{c}_k \varphi_k)^2\,dt \right)^{1/2} \\ &= o_P(1). \end{aligned}$$

Hence the result follows immediately from (3.28) and the diagonality of $\mathbf{Q}$. □

Proof of Theorem 3.6: By the ergodic theorem $a_i \xrightarrow{P} \langle \mu - \mu^*, \varphi_i \rangle\,,\ 1 \le i \le p$. Since relations (3.45) hold also under H_A, the result is proven. □

3.7 Bibliography

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley and Sons.

Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica 59*, 817–858.

Andrews, D. W. K. and J. C. Monahan (1992). An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica 60*, 953–966.

Aue, A., S. Hörmann, L. Horváth, M. Hušková, and J. Steinebach (2011). Sequential Testing for the Stability of Portfolio Betas. *Econometric Theory 00*, 000–000.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods.* Springer.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability.* Springer.

Doukhan, P. and S. Louhichi (1999). A New Weak Dependence and Applications to Moment Inequalities. *Stochastic Processes and their Applications 84*, 313–343.

Gabrys, R., L. Horváth, and P. Kokoszka (2010). Tests for Error Correlation in the Functional Linear Model. *Journal of the American Statistical Association 105*, 1113–1125.

Gabrys, R. and P. Kokoszka (2007). Portmanteau Test of Independence for Functional Observations. *Journal of the American Statistical Association 102*, 1338–1348.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press.

Hörmann, S. and P. Kokoszka (2010). Weakly Dependent Functional Data. *The Annals of Statistics 38*, 1845–1884.

Horváth, L., M. Hušková, and P. Kokoszka (2010). Testing the Stability of the Functional Autoregressive Process. *Journal of Multivariate Analysis 101*, 352–367.

Horváth, L. and P. Kokoszka (2011+). *Inference for Functional Data with Applications.* Springer Series in Statistics. Springer. Forthcoming.

Horváth, L., P. Kokoszka, and M. Reimherr (2009). Two Sample Inference in Functional Linear Models. *Canadian Journal of Statistics 37*, 571–591.

Kargin, V. and A. Onatski (2008). Curve Forecasting by Functional Autoregression. *Journal of Multivariate Analysis 99*, 2508–2526.

Lehmann, E. L. (1999). *Elements of Large Sample Theory.* Springer.

Müller, H.-G. and U. Stadtmüller (2005). Generalized Functional Linear Models. *The Annals of Statistics 33*, 774–805.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*, 703–08.

Panaretos, V. M., D. Kraus, and J. H. Maddocks (2010). Second-order Comparison of Gaussian Random Functions and the Geometry of DNA Minicircles. *Journal of the American Statistical Association 105*, 670–682.

Politis, D. N. and J. P. Romano (1996). On Flat-top Spectral Density Estimators for Homogeneous Random Fields. *Journal of Statistical Planning and Inference 51*, 41–53.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis.* Springer.

Wu, W. (2005). *Nonlinear System Theory: Another Look at Dependence*, Volume 102 of *Proceedings of The National Academy of Sciences of the United States.* National Academy of Sciences.

Wu, W. (2007). Strong Invariance Principles for Dependent Random Variables. *The Annals of Probability 35*, 2294–2320.

CHAPTER 4

A TEST OF SIGNIFICANCE IN FUNCTIONAL QUADRATIC REGRESSION[2]

We consider a quadratic functional regression model in which a scalar response depends on a functional predictor; the common functional linear model is a special case. We wish to test the significance of the nonlinear term in the model. We develop a testing method which is based on projecting the observations onto a suitably chosen finite dimensional space using functional principal component analysis. The asymptotic behavior of our testing procedure is established. A simulation study shows that the testing procedure has good size and power with finite sample sizes. We then apply our test to a data set provided by Tecator, which consists of near-infrared absorbance spectra and fat content of meat.

4.1 Introduction and results

In a predictive model, it may be more natural and appropriate for certain quantities to be represented as trajectories rather than a single number (Kirkpatrick and Heckman, 1989). For example, a young animal's size may be considered as a function of time, giving a growth trajectory. A model to predict a certain response from growth trajectories is useful to animal breeders because they may be able to produce more valuable animals by changing their growth patterns (Fitzhugh, 1976). Müller and Zhang (2005) used egg-laying trajectories from Mediterranean fruit flies to predict a female fly's remaining lifetime. Frank and Friedman (1993) and Wold (1993) provide an early discussion on the applications of principal components to analyze curves in chemistry. Yao and Müller (2010) and Borggaard and Thodberg (1992) used absorbance trajectories to predict the fat content of meat samples. The absorbance at any particular wavelength is a measurement related to the proportion of light that passes through a meat sample. A representative sample of 15 of the 240 absorbance trajectories are pictured in Figure 4.1.

[2]The content of this chapter is based on joint research with Lajos Horváth.

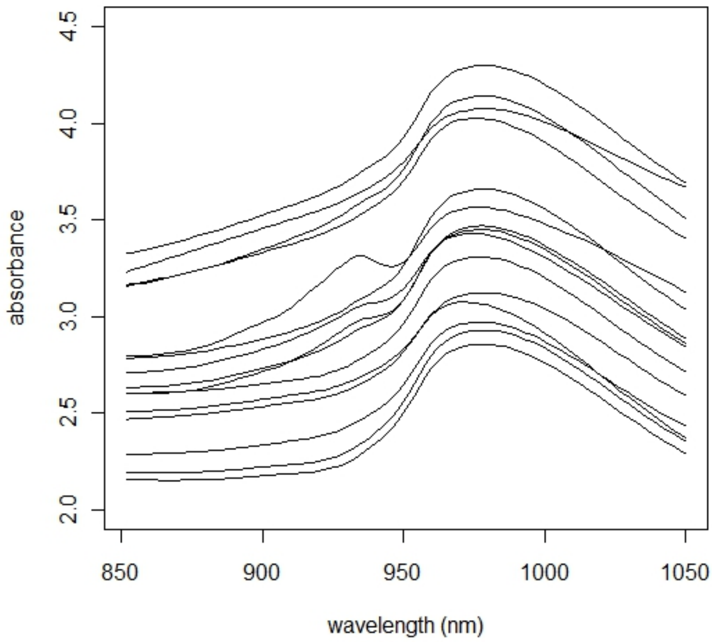


Figure 4.1. Absorbance trajectories from 15 samples of finely chopped pure meat.

In functional regression, special attention has been given to functional linear models (Cardot et al., 2003; Shen and Faraway, 2004; Cai and Hall, 2006; Hall and Horowitz, 2007). However, it is pointed out in Yao and Müller (2010) that this model imposes a constraint on the regression relationship that may not be appropriate in some scenarios. Yao and Müller (2010) generalized this to a functional polynomial model, which has greater flexibility. In functional polynomial regression, as in standard polynomial regression, one must balance the costs and benefits of using more parameters in the model. In this paper, we will develop a test to determine if a quadratic term is justified in the model or if a functional linear model adequately describes the regression relationship.

The functional quadratic model in which a scalar response, Y_n, is paired with a functional predictor, $X_n(t)$, is defined as

$$Y_n = \mu + \int_0^1 k(t) X_n^c(t)\, dt + \int_0^1 \int_0^1 h(s,t) X_n^c(s) X_n^c(t)\, dt\, ds + \varepsilon_n, \tag{4.1}$$

where $X_n^c(t) = X_n(t) - E\left(X_n(t)\right)$ is the centered predictor process. If $h(s,t) = 0$, then $\mu = E(Y_n)$ and (4.1) reduces to the functional linear model

$$Y_n = \mu + \int_0^1 k(t) X_n^c(t)\, dt + \varepsilon_n. \tag{4.2}$$

Cardot and Sarda (2011) and Mas and Pumo (2011) point out in their survey papers that since we can choose a function in (4.2), the functional linear model can be used in a large variety of applications. The functional linear model provides a very simple relation between $X_n(t)$ and Y_n, so it is important to check if the more involved quadratic model (4.1) provides a real improvement. In other words, one should test whether the quadratic term is really needed. To test the significance of the quadratic term in (4.1), we test the null hypothesis,

$$H_0 : h(s,t) = 0, \tag{4.3}$$

against the alternative

$$H_A : h(s,t) \neq 0.$$

To reduce the dimensionality and avoid overfitting in our functional regression model, we will project the predictor process onto a suitably chosen finite dimensional space. The space is spanned by the eigenfunctions of $C(t,s) = E(X_n(t) - \mu_X(t))(X_n(s) - \mu_X(s))$, the covariance function of the predictor process, where $\mu_X(t) = EX_n(t)$. We will denote the eigenfunctions and associated eigenvalues by $\{(v_i(t), \lambda_i),\ 1 \leq i \leq \infty\}$. We can and will assume that λ_i is the i^{th} largest eigenvalue and that the eigenfunctions are orthonormal. It is clear that we can assume that h is symmetric, and we also impose the condition that the kernels are in L^2:

$$h(s,t) = h(t,s) \text{ and } \int_0^1 \int_0^1 h^2(s,t)dtds < \infty, \tag{4.4}$$

$$\int_0^1 k^2(t)dt < \infty. \tag{4.5}$$

Thus we have the expansions

$$\begin{aligned} h(s,t) &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{i,j} v_j(s) v_i(t) \\ &= \sum_{i=1}^{\infty} a_{i,i} v_i(s) v_i(t) + \sum_{i=1}^{\infty} \sum_{j=i+1}^{\infty} a_{i,j} \left(v_j(s) v_i(t) + v_i(s) v_j(t) \right) \end{aligned} \tag{4.6}$$

and

$$k(t) = \sum_{i=1}^{\infty} b_i v_i(t). \tag{4.7}$$

By projecting onto the space spanned by $\{v_1, \ldots, v_p\}$ and using (4.6) and (4.7), we can write the model (4.1) as

$$Y_n = \mu + \sum_{i=1}^{p} b_i \langle X_n^c, v_i \rangle + \sum_{i=1}^{p} \sum_{j=i}^{p} (2 - 1\{i=j\}) a_{i,j} \langle X_n^c, v_i \rangle \langle X_n^c, v_j \rangle + \varepsilon_n^*, \tag{4.8}$$

where

$$\begin{aligned} \varepsilon_n^* = \varepsilon_n &+ \sum_{i=p+1}^{\infty} b_i \langle X_n^c, v_i \rangle + \sum_{i=p+1}^{\infty} \sum_{j=i}^{\infty} (2 - 1\{i=j\}) a_{i,j} \langle X_n^c, v_i \rangle \langle X_n^c, v_j \rangle \\ &+ \sum_{i=1}^{p} \sum_{j=p+1}^{\infty} 2 a_{i,j} \langle X_n^c, v_i \rangle \langle X_n^c, v_j \rangle. \end{aligned}$$

We note that (4.8) is written as a standard linear model, but the error term, ε_n^*, and the design points, $\{\langle X_n^c, v_i \rangle, 1 \le i \le p\}$, are dependent.

Unfortunately, we cannot use (4.8) directly for statistical inference since $v_i(t)$ and $\mu_X(t)$ are unknown. We estimate $\mu_X(t)$ and $C(t,s)$ with the corresponding empiricals

$$\bar{X}(t) = \frac{1}{N} \sum_{n=1}^{N} X_n(t)$$

and

$$\hat{C}(t,s) = \frac{1}{N} \sum_{n=1}^{N} \left(X_n(t) - \bar{X}(t) \right) \left(X_n(s) - \bar{X}(s) \right).$$

The eigenvalues and the corresponding eigenfunctions of $\hat{C}(t,s)$ are denoted by $\hat{\lambda}_1 \ge \hat{\lambda}_2 \ge \ldots$ and $\hat{v}_1, \hat{v}_2, \ldots$. Eigenfunctions corresponding to unique eigenvalues are uniquely determined up to signs. For this reason, we cannot expect more than to have $\hat{c}_i \hat{v}_i(t)$ be close to $v_i(t)$, where the $\hat{c}_i$'s are random signs. We replace equation (4.8) with

$$Y_n = \mu + \sum_{i=1}^{p} b_i \langle X_n - \bar{X}, \hat{c}_i \hat{v}_i \rangle + \sum_{i=1}^{p} \sum_{j=i}^{p} (2 - 1\{i = j\}) a_{i,j} \langle X_n - \bar{X}, \hat{c}_i \hat{v}_i \rangle \langle X_n - \bar{X}, \hat{c}_j \hat{v}_j \rangle + \varepsilon_n^{**}, \tag{4.9}$$

where

$$\begin{aligned} \varepsilon_n^{**} =& \varepsilon_n^* + \sum_{i=1}^{p} b_i \langle X_n^c, v_i - \hat{c}_i \hat{v}_i \rangle + \sum_{i=1}^{p} b_i \langle \bar{X} - \mu_X, \hat{c}_i \hat{v}_i \rangle \\ & - \sum_{i=1}^{p} \sum_{j=i}^{p} (2 - 1\{i = j\}) a_{i,j} \left(\langle X_n - \bar{X}, \hat{c}_i \hat{v}_i \rangle \langle X_n - \bar{X}, \hat{c}_j \hat{v}_j \rangle - \langle X_n^c, v_i \rangle \langle X_n^c, v_j \rangle \right). \end{aligned}$$

We can write (4.9) in the concise form

$$\mathbf{Y} = \hat{\mathbf{Z}} \begin{pmatrix} \tilde{\mathbf{A}} \\ \tilde{\mathbf{B}} \\ \mu \end{pmatrix} + \boldsymbol{\varepsilon}^{**}, \tag{4.10}$$

where

$$\begin{aligned} \mathbf{Y} &= \left(Y_1, Y_2, \ldots, Y_N \right)^T, & \tilde{\mathbf{A}} &= \operatorname{vech}\left(\{ \hat{c}_i \hat{c}_j a_{i,j} \left(2 - 1\{i = j\} \right), \ 1 \le i \le j \le p \}^T \right), \\ \tilde{\mathbf{B}} &= \left(\hat{c}_1 b_1, \hat{c}_2 b_2, \ldots, \hat{c}_p b_p \right)^T, & \boldsymbol{\varepsilon}^{**} &= \left(\varepsilon_1^{**}, \varepsilon_2^{**}, \ldots, \varepsilon_N^{**} \right)^T, \end{aligned}$$

and

$$\hat{\mathbf{Z}} = \begin{pmatrix} \hat{\mathbf{D}}_1^T & \hat{\mathbf{F}}_1^T & 1 \\ \hat{\mathbf{D}}_2^T & \hat{\mathbf{F}}_2^T & 1 \\ \vdots & \vdots & \vdots \\ \hat{\mathbf{D}}_N^T & \hat{\mathbf{F}}_N^T & 1 \end{pmatrix}$$

with

$$\begin{aligned} \hat{\mathbf{D}}_n &= \operatorname{vech}\left(\{ \langle \hat{v}_i, X_n - \bar{X} \rangle \langle \hat{v}_j, X_n - \bar{X} \rangle, \ 1 \le i \le j \le p \}^T \right), \\ \hat{\mathbf{F}}_n &= \left(\langle X_n - \bar{X}, \hat{v}_1 \rangle, \quad \langle X_n - \bar{X}, \hat{v}_2 \rangle, \quad \ldots, \quad \langle X_n - \bar{X}, \hat{v}_p \rangle \right)^T. \end{aligned}$$

The half-vectorization, vech$(\cdot)$, stacks the columns of the lower triangular portion of the matrix under each other. Although we write our model in the form of a general linear model, it is important to note that it is not a classical linear model. First, $\boldsymbol{\varepsilon}^{**}$ is correlated with $\hat{\mathbf{Z}}$ because $\boldsymbol{\varepsilon}^{**}$ contains additional error terms which come from projecting onto a p-dimensional space. Another important difference between (4.10) and a classical linear model is that the parameters to be estimated, $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$, are random; they depend on the random signs, $\hat{c}_i$. We estimate $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and μ using the least squares estimator:

$$\begin{pmatrix} \hat{\mathbf{A}} \\ \hat{\mathbf{B}} \\ \hat{\mu} \end{pmatrix} = \left(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}^T\mathbf{Y}. \tag{4.11}$$

To represent elements of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$, we will use the notation that $\hat{\mathbf{A}} = \text{vech}(\{\hat{a}_{i,j}(2 - 1\{i = j\}), 1 \leq i \leq j \leq p\}^T)$ and $\hat{\mathbf{B}} = \left(\hat{b}_1, \quad \hat{b}_2, \quad \ldots, \quad \hat{b}_p\right)^T$.

We expect, under H_0, that $\hat{\mathbf{A}}$ will be close to zero since $\tilde{\mathbf{A}}$ is zero. If H_0 is not correct, we expect the magnitude of $\hat{\mathbf{A}}$ to be relatively large. This suggests that a testing procedure could be based on $\hat{\mathbf{A}}$. Due to the random signs coming from the estimation of the eigenfunctions, $\hat{\mathbf{A}}$ will not be asymptotically normal. However, if the random signs are "taken out," asymptotic normality can be established. Hence our test statistic will be a quadratic form of $\hat{\mathbf{A}}$ with some random weight matrices. Let

$$\hat{\mathbf{G}} = \frac{1}{N}\sum_{n=1}^{N}\hat{\mathbf{D}}_n\hat{\mathbf{D}}_n^T,$$

$$\hat{\mathbf{M}} = \frac{1}{N}\sum_{n=1}^{N}\hat{\mathbf{D}}_n,$$

and

$$\hat{\tau}^2 = \frac{1}{N}\sum_{n=1}^{N}\hat{\varepsilon}_n^2,$$

where

$$\hat{\varepsilon}_n = Y_n - \hat{\mu} - \sum_{i=1}^{p}\hat{b}_i\langle X_n - \bar{X}, \hat{v}_i\rangle - \sum_{i=1}^{p}\sum_{j=i}^{p}(2 - 1\{i = j\})\hat{a}_{i,j}\langle X_n - \bar{X}, \hat{v}_i\rangle\langle X_n - \bar{X}, \hat{v}_j\rangle$$

are the residuals under H_0. We reject the null hypothesis if

$$U_N = \frac{N}{\hat{\tau}^2}\hat{\mathbf{A}}^T(\hat{\mathbf{G}} - \hat{\mathbf{M}}\hat{\mathbf{M}}^T)\hat{\mathbf{A}}$$

is large. The main result of this paper is the asymptotic distribution of U_N under the null hypothesis. First, we discuss the assumptions needed to establish asymptotics for U_N:

Assumption 4.1. $\{X_n(t), n \geq 1\}$ *is a sequence of independent, identically distributed Gaussian processes.*

Assumption 4.2.

$$E\left(\int_0^1 X_n^2(t)\,dt\right)^4 < \infty.$$

Assumption 4.3. *$\{\varepsilon_n\}$ is a sequence of independent, identically distributed random variables satisfying $E\varepsilon_n = 0$ and $E\varepsilon_n^4 < \infty$,*

and

Assumption 4.4. *the sequences $\{\varepsilon_n\}$ and $\{X_n(t)\}$ are independent.*

The last condition is standard in functional data analysis. It implies that the eigenfunctions $v_1, v_2, \ldots, v_p$ are unique up to a sign.

Assumption 4.5.

$$\lambda_1 > \lambda_2 > \cdots > \lambda_{p+1}.$$

Theorem 4.1. *If H_0, (4.5) and Assumptions 4.1–4.5 are satisfied, then*

$$U_N \xrightarrow{d} \chi^2(r),$$

where $r = p(p+1)/2$ is the dimension of the vector $\hat{\mathbf{A}}$.

The proof of Theorem 4.1 is given in Section 4.4.

Remark 4.1. *By the Karhunen-Loève expansion, every centered, square integrable process, $X_n^c(t)$, can be written as*

$$X_n^c(t) = \sum_{\ell=1}^{\infty} \xi_{n,\ell}\varphi_\ell(t),$$

where φ_ℓ are orthonormal functions. Assumption 4.1 can be replaced with the requirement that $\xi_{n,1}$, $\xi_{n,2}$, $\ldots$, $\xi_{n,p}$ are independent with $E\xi_{n,\ell}^3 = 0$ and $E\xi_{n,\ell} = 0$ for all $1 \le \ell \le p$.

Our last result provides a simple condition for the consistency of the test based on U_N. Let $\mathbf{A} = \text{vech}(\{a_{i,j}(2 - 1\{i = j\}),\ 1 \le i \le j \le p\}^T)$, i.e. the first $r = p(p+1)/2$ coefficients in the expansion of h in (4.6).

Theorem 4.2. *If* (4.4), (4.5), *Assumptions 4.1–4.5 are satisfied and $\mathbf{A} \neq \mathbf{0}$, then we have that*

$$U_N \xrightarrow{P} \infty.$$

The condition $\mathbf{A} \neq \mathbf{0}$ means that h is not the 0 function in the space spanned by the functions $v_i(t)v_j(s), 1 \le i, j \le p$.

4.2 A simulation study

In this section, we investigate the empirical size and power of the testing procedure for finite sample sizes. Seeking to obtain a test of size $\alpha = .01$, $.05$, or $.10$, a rejection region was chosen according to the limiting distribution of the test statistic. Since the limiting distribution is $\chi^2(r)$, the rejection region is (Δ, ∞), where $P\left(\chi^2(r) > \Delta\right) = \alpha$. Simulated data was then used to compute the outcome of the test statistic. Iterating this procedure 2,000 times, we kept track of the proportion of times that the outcome fell in the predetermined rejection region. When simulations are done under H_0, this gives us the empirical size of the test, which we expect to be close to the nominal size, α, for large sample sizes. When simulations are done under the alternative, H_A, the proportion gives us the empirical power of the test.

In our first simulation study, the ε_n's were generated according to the distribution of independent standard normals. We generated the $X_n(t)$'s according to the distribution of independent standard Brownian motions. Then, using $k(t) = 1$ and $h(s,t) = c$, we obtained Y_n according to (4.1). Thus the power of the test is a function of the parameter c. In particular, when $c = 0$, the null hypothesis is true. The resulting empirical size and power are given in Table 4.1.

The distribution of our test statistic has been shown to converge to a $\chi^2(r)$. Thus we expect the empirical and nominal size to be close for samples of size $N = 200$ and even closer when $N = 500$, as observed in Table 4.1. Since our testing procedure depends on the choice of how many principal components to keep, results are given in Table 4.1 for $p = 1$, 2, and 3. One possible method of selecting p is to follow the advice of Ramsay and Silverman (2005) and choose p so that approximately 85% of the variance within a sample is described by the first p principal components.

Although Theorem 4.1 is proven under the assumption that $X_n(t)$ is a Gaussian process, the result of Theorem 4.1 holds under relaxed conditions as discussed in Remark 4.1. We will now investigate the empirical size and power of our test when $X_n(t)$ is not a Gaussian process. We generate the ε_n's according to a uniform distribution on $(-0.5, 0.5)$. The predictors, $X_n(t)$, are generated according to

$$X_n(t) = \left(T_{1,n} + T_{2,n}t + T_{3,n}(2t^2 - 1) + T_{4,n}(4t^3 - 3t)\right)/4,$$

where $\{T_{i,n}, 1 \le i \le 4, 1 \le n\}$ are iid random variables having a t-distribution with 5 degrees of freedom. The polynomials in the definition of $X_n(t)$ are the orthogonal Chebyshev polynomials. The resulting empirical size and power are given in Table 4.2. We see

Table 4.1. Empirical power of test (in %) based on 2,000 simulations using iid Brownian motions for $X_n(t)$ and iid standard normals for ε_n.

c	$\alpha = .01$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	1.02	1.37	1.95	1.10	1.30	1.15
0.2	10.81	6.87	6.52	30.35	20.35	12.85
0.4	49.51	37.24	29.76	91.90	84.25	74.35
0.6	86.68	77.74	70.19	100.00	99.70	98.75
0.8	98.50	96.05	92.98	100.00	100.00	100.00
1.0	99.94	99.57	99.05	100.00	100.00	100.00

c	$\alpha = .05$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	5.15	6.00	7.44	5.60	5.75	6.05
0.2	25.90	19.17	18.02	53.05	40.00	31.35
0.4	72.10	60.31	50.38	97.90	93.70	88.55
0.6	95.21	90.43	85.77	100.00	99.90	99.60
0.8	99.60	98.90	97.60	100.00	100.00	100.00
1.0	99.99	99.87	99.84	100.00	100.00	100.00

c	$\alpha = .10$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	10.27	11.18	13.35	10.60	11.05	11.55
0.2	36.60	29.50	27.03	65.00	52.45	43.75
0.4	80.89	71.08	62.27	99.30	96.60	93.10
0.6	97.60	94.77	90.91	100.00	99.95	99.75
0.8	99.85	99.47	98.57	100.00	100.00	100.00
1.0	99.99	99.95	99.91	100.00	100.00	100.00

from Table 4.2 that our testing procedure is robust against non-Gaussian observations. Comparing Tables 4.1 and 4.2, we see that the value of the test statistics tends to be larger if the X_n's are not normally distributed for small N. The overrejection fades as N gets larger so in case of non-Gaussian X_n's, larger sample sizes are needed. This also explains the somewhat better power of the procedure in the case of non-Gaussian errors.

4.3 Application to spectral data

In this section we apply our test to the data set collected by Tecator and available at http://lib.stat.cmu.edu/datasets/tecator. Tecator used 240 samples of finely chopped pure meat with different fat contents. For each sample of meat, a 100 channel spectrum of absorbances was recorded using a Tecator Infratec food and feed analyzer. These absorbances can be thought of as a discrete approximation to the continuous record, $X_n(t)$. Also, for each sample of meat, the fat content, Y_n was measured by analytic chemistry.

The absorbance curve measured from the n^{th} meat sample is given by $X_n(t) = \log_{10}(I_0/I)$, where t is the wavelength of the light, I_0 is the intensity of the light before passing through the meat sample, and I is the intensity of the light after it passes through the meat sample. The Tecator Infratec food and feed analyzer measured absorbance at 100 different wavelengths between 850 and 1050 nanometers. This gives the values of $X_n(t)$ on a discrete grid from which we can use cubic splines to interpolate the values anywhere within the interval. A representative sample of 15 of the 240 absorbance trajectories are pictured in Figure 4.1.

Yao and Müller (2010) proposed using a functional quadratic model to predict the fat content, Y_n, of a meat sample based on its absorbance spectrum, $X_n(t)$. We are interested in determining whether the quadratic term in (4.1) is needed by testing its significance for this data set. From the data, we calculate U_{240}. The p-value is then $P\left(\chi^2(r) > U_{240}\right)$. The test statistic and hence the p-value are influenced by the number of principal components that we choose to keep. If we select p according to the advice of Ramsay and Silverman (2005), we will keep only $p = 1$ principal component because this explains more than 85% of the variation between absorbance curves in the sample. Table 4.3 gives p-values obtained using $p = 1$, 2, and 3 principal components, which strongly supports that the quadratic regression provides a better model for the Tecator data.

4.4 Proof of Theorem 4.1

Proof of Theorem 4.1. We have from (4.10) and (4.11) that

Table 4.2. Empirical power of test (in %) based on 2,000 simulations using non-Gaussian $X_n(t)$ and non-normal ε_n.

c	$\alpha = .01$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	2.40	1.20	1.85	1.75	1.45	1.35
0.2	57.70	46.75	37.50	93.75	90.30	82.55
0.4	96.90	95.55	91.20	100.00	100.00	100.00
0.6	99.90	100.00	99.70	100.00	100.00	100.00
0.8	100.00	100.00	100.00	100.00	100.00	100.00
1.0	100.00	100.00	100.00	100.00	100.00	100.00
c	$\alpha = .05$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	8.00	5.75	8.15	7.20	5.45	6.10
0.2	74.50	64.55	56.45	98.55	96.30	92.00
0.4	99.40	98.35	96.55	100.00	100.00	100.00
0.6	99.95	100.00	99.85	100.00	100.00	100.00
0.8	100.00	100.00	100.00	100.00	100.00	100.00
1.0	100.00	100.00	100.00	100.00	100.00	100.00
c	$\alpha = .10$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	13.60	12.15	14.60	13.60	10.35	11.50
0.2	82.30	74.25	65.55	98.90	97.70	95.25
0.4	99.65	99.10	97.95	100.00	100.00	100.00
0.6	99.95	100.00	99.90	100.00	100.00	100.00
0.8	100.00	100.00	100.00	100.00	100.00	100.00
1.0	100.00	100.00	100.00	100.00	100.00	100.00

Table 4.3. p-values (in %) obtained by applying our testing procedure to the Tecator data set with $p = 1$, 2, and 3 principal components.

p	1	2	3
p-value	1.25	13.15	0.00

$$
\begin{aligned}
\begin{pmatrix} \hat{\mathbf{A}} \\ \hat{\mathbf{B}} \\ \hat{\mu} \end{pmatrix} &= \left(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}^T\left(\hat{\mathbf{Z}}\begin{pmatrix} \tilde{\mathbf{A}} \\ \tilde{\mathbf{B}} \\ \mu \end{pmatrix} + \boldsymbol{\varepsilon}^{**}\right) \\
&= \begin{pmatrix} \tilde{\mathbf{A}} \\ \tilde{\mathbf{B}} \\ \mu \end{pmatrix} + \left(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**}.
\end{aligned} \tag{4.12}
$$

We also note that, under the null hypothesis, $a_{i,j} = 0$ for all i and j and therefore ε_n^* and ε_n^{**} of (4.8) and (4.9) reduce to

$$
\varepsilon_n^* = \varepsilon_n + \sum_{i=p+1}^{\infty} b_i \langle X_n^c, v_i \rangle
$$

and

$$
\varepsilon_n^{**} = \varepsilon_n^* + \sum_{i=1}^{p} b_i \langle X_n^c, v_i - \hat{c}_i\hat{v}_i \rangle + \sum_{i=1}^{p} b_i \langle \bar{X} - \mu_X, \hat{c}_i\hat{v}_i \rangle.
$$

To obtain the limiting distribution of $\sqrt{N}\hat{\mathbf{A}}$, we need to consider the vector $\sqrt{N}\left(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**}$. We will show in Lemmas 4.2–4.7 that

$$
\left(\left(\frac{\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}}{N}\right) - \begin{pmatrix} \boldsymbol{\zeta}\mathbf{G}\boldsymbol{\zeta} & \mathbf{0}_{r\times p} & \mathbf{M} \\ \mathbf{0}_{p\times r} & \boldsymbol{\Lambda} & \mathbf{0}_{p\times 1} \\ \mathbf{M}^T & \mathbf{0}_{1\times p} & 1 \end{pmatrix}\right) = o_P(1), \tag{4.13}
$$

where $\boldsymbol{\zeta}$ is an unobservable matrix of random signs, $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$, $\mathbf{M} = E(\mathbf{D}_n)$, and $\mathbf{G} = E\left(\mathbf{D}_n\mathbf{D}_n^T\right)$, where

$$
\mathbf{D}_n = \operatorname{vech}\left(\{\langle v_i, X_n^c\rangle\langle v_j, X_n^c\rangle,\ 1 \le i \le j \le p\}^T\right).
$$

We see from (4.13) that the vector $\sqrt{N}\left(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**}$ has the same limiting distribution as

$$
\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\varepsilon_n^{**}\begin{pmatrix} \boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta} & \mathbf{0}_{r\times p} & -\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M} \\ \mathbf{0}_{p\times r} & \boldsymbol{\Lambda}^{-1} & \mathbf{0}_{p\times 1} \\ -\mathbf{M}^T\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1} & \mathbf{0}_{1\times p} & 1+\mathbf{M}^T\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\mathbf{M} \end{pmatrix}\begin{pmatrix} \hat{\mathbf{D}}_n \\ \hat{\mathbf{F}}_n \\ 1 \end{pmatrix}. \tag{4.14}
$$

Since we are interested only in $\sqrt{N}\hat{\mathbf{A}}$ we need consider only the first $r = p(p+1)/2$ elements of the vector in (4.14). In Lemma 4.8 we show that these are given by

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\varepsilon_n^{**}\begin{pmatrix}\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta} & \mathbf{0}_{r\times p} & -\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M}\end{pmatrix}\begin{pmatrix}\hat{\mathbf{D}}_n\\ \hat{\mathbf{F}}_n\\ 1\end{pmatrix}$$
$$=\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\hat{\mathbf{D}}_n-\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M}\right)$$
$$=\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\varepsilon_n^{**}\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\left(\hat{\mathbf{D}}_n-\mathbf{M}\right).$$

Then, in Lemma 4.9 we prove that

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\left(\hat{\mathbf{D}}_n-\mathbf{M}\right)\xrightarrow{d}N\left(0,\tau^2\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\right),$$

where $\tau^2=\mathrm{Var}\left(\varepsilon_1^*\right)$. Finally, in Lemmas 4.10 and 4.11, we show that $\hat{\tau}^2-\tau^2=o_P(1)$. As a consequence of (4.13), we see that $\left(\hat{\mathbf{G}}-\hat{\mathbf{M}}\hat{\mathbf{M}}^T\right)-\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)\boldsymbol{\zeta}=o_P(1)$. Since $\boldsymbol{\zeta}$ is a diagonal matrix of signs, $\boldsymbol{\zeta}\boldsymbol{\zeta}=I$, completing the proof of Theorem 4.1. □

4.5 Proof of Theorem 4.2

We provide only an outline of the proof since it follows the arguments used in the proof of Theorem 4.1. However, the arguments are simple since instead of obtaining an asymptotic limit distribution we only establish the weak law

$$\hat{\mathbf{A}}^T(\hat{\mathbf{G}}-\hat{\mathbf{M}}\hat{\mathbf{M}}^T)\hat{\mathbf{A}}\ \xrightarrow{P}\ \mathbf{A}^T(\mathbf{G}-\mathbf{M}\mathbf{M}^T)\mathbf{A},\tag{4.15}$$

where $\mathbf{A}=\mathrm{vech}\left(\{a_{i,j}\left(2-1\{i=j\}\right),\ 1\le i\le j\le p\}^T\right)$ is like the vector $\tilde{\mathbf{A}}$ except without the random signs.

First we note that according to Lemma 4.1, the estimation of $v_1,\dots,v_p$ by $\hat{v}_1,\dots,\hat{v}_p$ causes only the introduction of the random signs $\hat{c}_1,\dots,\hat{c}_p$. As in the proof of Theorem 4.1 one can verify that

$$\hat{\mathbf{A}}-\boldsymbol{\zeta}\mathbf{A}\ \xrightarrow{P}\ \mathbf{0}.$$

Lemmas 4.2 and 4.6 hold under H_0 as well as under H_A. This gives

$$\hat{\mathbf{G}}-\boldsymbol{\zeta}\mathbf{G}\boldsymbol{\zeta}=o_P(1)$$

and

$$\hat{\mathbf{M}}\hat{\mathbf{M}}^T-\boldsymbol{\zeta}\mathbf{M}\mathbf{M}^T\boldsymbol{\zeta}=o_P(1),$$

completing the proof of (4.15).

4.6 Technical lemmas

Throughout the proofs in this section we will use $\|\cdot\|_1$ to be the 1-norm and $\|\cdot\|_2$ to be 2-norm on the unit interval, square, cube, or hypercube. The null hypothesis, H_0, is assumed throughout this section. We will make frequent use of the following lemma, which is established in Dauxois et al. (1982) and Bosq (2000).

Lemma 4.1. *If Assumptions 4.1, 4.2, and 4.5 hold, then*

$$\|\hat{c}_i \hat{v}_i(t) - v_i(t)\| = O_P\left(N^{-1/2}\right)$$

for each $1 \leq i \leq p$.

Lemma 4.2. *If Assumptions 4.1, 4.2, and 4.5 hold, then there is a non-random matrix* $\mathbf{G}$ *such that*

$$\left(\hat{\mathbf{G}} - \boldsymbol{\zeta}\mathbf{G}\boldsymbol{\zeta}\right) = o_P\left(1\right),$$

where $\hat{\mathbf{G}} = N^{-1}\sum_{n=1}^{N} \hat{\mathbf{D}}_n \hat{\mathbf{D}}_n^T$ *and* $\boldsymbol{\zeta} = \operatorname{diag}\left(\operatorname{vech}(\{\hat{c}_i\hat{c}_j, 1 \leq i \leq j \leq p\}^T)\right)$.

Proof. By the Karhunen-Loéve expansion we have

$$X_n^c(t) = \sum_{\ell=1}^{\infty} \lambda_\ell^{1/2} \xi_\ell^{(n)} v_\ell(t). \tag{4.16}$$

Therefore an element of $\mathbf{D}_n\mathbf{D}_n^T$ is of the form $\sqrt{\lambda_i\lambda_j\lambda_k\lambda_\ell}\,\xi_i^{(n)}\xi_j^{(n)}\xi_k^{(n)}\xi_\ell^{(n)}$. Hence using the strong law of large numbers we conclude

$$\frac{1}{N}\sum_{n=1}^{N} \mathbf{D}_n\mathbf{D}_n^T \overset{a.s.}{\to} \mathbf{G},$$

where $\mathbf{G} = E\left(\mathbf{D}_n\mathbf{D}_n^T\right)$. Thus it suffices to show that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\boldsymbol{\zeta}\hat{\mathbf{D}}_n\hat{\mathbf{D}}_n^T\boldsymbol{\zeta} - \mathbf{D}_n\mathbf{D}_n^T\right) = o_P\left(1\right). \tag{4.17}$$

Expressing (4.17) elementwise, we obtain

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}\Big(&\langle X_n - \bar{X}, \hat{c}_i\hat{v}_i\rangle\langle X_n - \bar{X}, \hat{c}_j\hat{v}_j\rangle\langle X_n - \bar{X}, \hat{c}_k\hat{v}_k\rangle\langle X_n - \bar{X}, \hat{c}_\ell\hat{v}_\ell\rangle \\ &- \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\langle X_n^c, v_k\rangle\langle X_n^c, v_\ell\rangle\Big) = o_P\left(1\right).\end{aligned} \tag{4.18}$$

In order to prove (4.18), it is enough to show that

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}\Big(&\langle X_n^c, \hat{c}_i\hat{v}_i\rangle\langle X_n^c, \hat{c}_j\hat{v}_j\rangle\langle X_n^c, \hat{c}_k\hat{v}_k\rangle\langle X_n^c, \hat{c}_\ell\hat{v}_\ell\rangle \\ &- \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\langle X_n^c, v_k\rangle\langle X_n^c, v_\ell\rangle\Big) = o_P\left(1\right)\end{aligned} \tag{4.19}$$

and

$$\frac{1}{N}\sum_{n=1}^{N}\Big(\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle X_n-\bar{X},\hat{c}_j\hat{v}_j\rangle\langle X_n-\bar{X},\hat{c}_k\hat{v}_k\rangle\langle X_n-\bar{X},\hat{c}_\ell\hat{v}_\ell\rangle \\ -\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\langle X_n^c,\hat{c}_j\hat{v}_j\rangle\langle X_n^c,\hat{c}_k\hat{v}_k\rangle\langle X_n^c,\hat{c}_\ell\hat{v}_\ell\rangle\Big)=o_P(1)\,. \tag{4.20}$$

We only establish (4.19), since the proof of (4.20) is essentially the same. Using Hölder's inequality, we obtain

$$\Bigg|\int_0^1\int_0^1\int_0^1\int_0^1\left(\frac{1}{N}\sum_{n=1}^{N}X_n^c(s)X_n^c(t)X_n^c(u)X_n^c(w)\right) \\ \times\left(\hat{c}_i\hat{v}_i(s)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\hat{c}_\ell\hat{v}_\ell(w)-v_i(s)v_j(t)v_k(u)v_\ell(w)\right)\,ds\,dt\,du\,dw\Bigg| \\ \le\left\|\frac{1}{N}\sum_{n=1}^{N}X_n^c(s)X_n^c(t)X_n^c(u)X_n^c(w)\right\|_2 \\ \times\left\|\hat{c}_i\hat{v}_i(s)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\hat{c}_\ell\hat{v}_\ell(w)-v_i(s)v_j(t)v_k(u)v_\ell(w)\right\|_2.$$

By the law of large numbers in Hilbert spaces (cf. (Bosq, 2000)), we have that

$$\left\|\frac{1}{N}\sum_{n=1}^{N}X_n^c(s)X_n^c(t)X_n^c(u)X_n^c(w)\right\|_2=O_P(1)\,,$$

so it remains only to show that

$$\left\|\hat{c}_i\hat{v}_i(s)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\hat{c}_\ell\hat{v}_\ell(w)-v_i(s)v_j(t)v_k(u)v_\ell(w)\right\|_2=o_P(1)\,.$$

Using Minkowski's inequality, Fubini's Theorem, the fact that $\|\hat{v}_i\|_2=\|v_i\|_2=1$, and then Lemma 4.1, we obtain

$$\begin{aligned}
&\left\|\hat{c}_i\hat{v}_i(s)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\hat{c}_\ell\hat{v}_\ell(w)-v_i(s)v_j(t)v_k(u)v_\ell(w)\right\|_2\\
&\quad\le\left\|\left(\hat{c}_i\hat{v}_i(s)-v_i(s)\right)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\hat{c}_\ell\hat{v}_\ell(w)\right\|_2\\
&\qquad+\left\|v_i(s)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\left(\hat{c}_\ell\hat{v}_\ell(w)-v_\ell(w)\right)\right\|_2\\
&\qquad+\left\|v_i(s)\hat{c}_j\hat{v}_j(t)\left(\hat{c}_k\hat{v}_k(u)-v_k(u)\right)v_\ell(w)\right\|_2\\
&\qquad+\left\|v_i(s)\left(\hat{c}_j\hat{v}_j(t)-v_j(t)\right)v_k(u)v_\ell(w)\right\|_2\\
&\quad=\left\|\hat{c}_i\hat{v}_i-v_i\right\|_2+\left\|\hat{c}_j\hat{v}_j-v_j\right\|_2+\left\|\hat{c}_k\hat{v}_k-v_k\right\|_2+\left\|\hat{c}_\ell\hat{v}_\ell-v_\ell\right\|_2\\
&\quad=O_P\left(N^{-1/2}\right).
\end{aligned}$$

Hence (4.19) is proven which also completes the proof of Lemma 4.2. □

Lemma 4.3. *If Assumptions 4.1, 4.2, and 4.5 hold, then*

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\mathbf{F}}_n\hat{\mathbf{D}}_n^T=o_P(1)\,.$$

Proof. We see from (4.16) that an element of $\mathbf{F}_n\mathbf{D}_n^T$ has the form $\sqrt{\lambda_i\lambda_j\lambda_k}\xi_i^{(n)}\xi_j^{(n)}\xi_k^{(n)}$, where $\mathbf{F}_n = \left(\langle X_n^c, v_1\rangle, \quad \langle X_n^c, v_2\rangle, \quad \dots, \quad \langle X_n^c, v_p\rangle\right)^T$. We observe that $E\xi_i^{(n)}\xi_j^{(n)}\xi_k^{(n)} = 0$, so using the central limit theorem, we have

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{F}_n\mathbf{D}_n^T = O_P\left(N^{-1/2}\right).$$

Repeating the arguments in the proof (4.18), one can verify that

$$\frac{1}{N}\sum_{n=1}^{N}\Big(\langle X_n - \bar{X}, \hat{c}_i\hat{v}_i\rangle\langle X_n - \bar{X}, \hat{c}_j\hat{v}_j\rangle\langle X_n - \bar{X}, \hat{c}_k\hat{v}_k\rangle \\ - \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\langle X_n^c, v_k\rangle\Big) = o_P(1). \tag{4.21}$$

Since random signs do not affect convergence to zero, the proof is complete. □

Lemma 4.4. *If Assumptions 4.1, 4.2, and 4.5 hold, then*

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\mathbf{F}}_n\hat{\mathbf{F}}_n^T - \mathbf{\Lambda} = o_P(1),$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$.

Proof. By (4.16), an element of $\mathbf{F}_n\mathbf{F}_n^T$ is of the form $\sqrt{\lambda_i\lambda_j}\xi_i^{(n)}\xi_j^{(n)}$. Since $E\xi_i^{(n)}\xi_j^{(n)} = 1\{i = j\}$, according to the law of large numbers we have

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{F}_n\mathbf{F}_n^T - \mathbf{\Lambda} = o_P(1).$$

Thus it suffices to demonstrate that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n - \bar{X}, \hat{v}_i\rangle\langle X_n - \bar{X}, \hat{v}_j\rangle - \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\right) = o_P(1). \tag{4.22}$$

Since random signs do not affect convergence to zero, multiplying $\hat{v}_i$ by $\hat{c}_i$ and $\hat{v}_j$ by $\hat{c}_j$ will not affect convergence when $i \neq j$. If $i = j$, then $\hat{c}_i\hat{c}_j = \hat{c}_i^2 = 1$. Therefore, it suffices to show that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n - \bar{X}, \hat{c}_i\hat{v}_i\rangle\langle X_n - \bar{X}, \hat{c}_j\hat{v}_j\rangle - \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\right) = o_P(1). \tag{4.23}$$

One can show (4.23) in exactly the same way we established (4.18) in the proof of Lemma 4.2. This completes the proof. □

Lemma 4.5. *If Assumptions 4.1, 4.2, and 4.5 hold, then*

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\mathbf{F}}_n = o_P\left(1\right).$$

Proof. Using (4.16), an element of $\mathbf{F}_n$ has the form $\sqrt{\lambda_i}\xi_i^{(n)}$, so the law of large numbers implies that

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{F}_n = o_P\left(1\right).$$

The proof will be completed by establishing that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{F}_n - \hat{\mathbf{F}}_n\right) = o_P\left(1\right). \tag{4.24}$$

We express (4.24) componentwise and obtain

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, v_i\rangle - \langle X_n - \bar{X}, \hat{v}_i\rangle\right) = o_P\left(1\right). \tag{4.25}$$

Since random signs do not affect convergence to zero, it suffices to show that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, v_i\rangle - \langle X_n - \bar{X}, \hat{c}_i\hat{v}_i\rangle\right) = o_P\left(1\right). \tag{4.26}$$

We will establish (4.26) in two steps. We will show that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, v_i\rangle - \langle X_n^c, \hat{c}_i\hat{v}_i\rangle\right) = o_P\left(1\right). \tag{4.27}$$

Then, we will establish that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, \hat{c}_i\hat{v}_i\rangle - \langle X_n - \bar{X}, \hat{c}_i\hat{v}_i\rangle\right) = o_P\left(1\right). \tag{4.28}$$

Using the central limit theorem in Hilbert spaces with Lemma 4.1 we conclude

$$\begin{aligned}\left|\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, v_i\rangle - \langle X_n^c, \hat{c}_i\hat{v}_i\rangle\right)\right| &\leq \left|\left|\frac{1}{N}\sum_{n=1}^{N}X_n^c(t)\left(v_i - \hat{c}_i\hat{v}_i\right)\right|\right|_1 \\ &\leq \left|\left|\frac{1}{N}\sum_{n=1}^{N}X_n^c(t)\right|\right|_2 ||v_i - \hat{c}_i\hat{v}_i||_2 \\ &= O_P\left(N^{-1}\right),\end{aligned}$$

and by the same arguments we have

$$\begin{aligned}\left|\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, \hat{c}_i\hat{v}_i\rangle - \langle X_n - \bar{X}, \hat{c}_i\hat{v}_i\rangle\right)\right| &= \left|\langle \mu_X - \bar{X}, \hat{c}_i\hat{v}_i\rangle\right| \\ &\leq \left|\left|\left(\mu_X(t) - \bar{X}(t)\right)\hat{c}_i\hat{v}_i(t)\right|\right|_1 \\ &\leq \left|\left|\mu_X(t) - \bar{X}(t)\right|\right|_2 \\ &= o_P\left(1\right).\end{aligned}$$

$\square$

Lemma 4.6. *If Assumptions 4.1, 4.2, and 4.5 hold, then*

$$\hat{\mathbf{M}} - \mathbf{M} = o_P(1).$$

where $\hat{\mathbf{M}} = N^{-1}\sum_{n=1}^{N}\hat{\mathbf{D}}_n$ *and* $\mathbf{M} = E(\mathbf{D}_n)$.

Proof. An arbitrary element of $\hat{\mathbf{D}}_n$ is of the form

$$\frac{1}{N}\sum_{n=1}^{N}\langle X_n - \bar{X}, \hat{v}_i\rangle\langle X_n - \bar{X}, \hat{v}_j\rangle.$$

Since this is exactly the same as the form of an arbitrary element of $\hat{\mathbf{F}}_n\hat{\mathbf{F}}_n^T$, Lemma 4.6 follows from the proof of Lemma 4.4. Note in particular that when $i \neq j$, the sum converges to zero and is unaffected by signs, and when $i = j$, the signs cancel each other out. For this reason, $\boldsymbol{\zeta}\mathbf{M} = \mathbf{M}$, rendering it unnecessary to multiply $\mathbf{M}$ by $\boldsymbol{\zeta}$ in the statement of the lemma. $\square$

Lemma 4.7. *If Assumptions 4.1, 4.2, and 4.5 hold, then*

$$\left(\left(\frac{\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}}{N}\right) - \begin{pmatrix} \boldsymbol{\zeta}\mathbf{G}\boldsymbol{\zeta} & \mathbf{0}_{r\times p} & \mathbf{M} \\ \mathbf{0}_{p\times r} & \boldsymbol{\Lambda} & \mathbf{0}_{p\times 1} \\ \mathbf{M}^T & \mathbf{0}_{1\times p} & 1 \end{pmatrix}\right) = o_P(1).$$

Proof. This follows immediately from Lemmas 4.2–4.6. $\square$

We will now use Lemma 4.7 to separate our estimate, $\hat{\mathbf{A}}$, of $\tilde{\mathbf{A}}$ from the estimates of the other parameters in (4.11).

Lemma 4.8. *If Assumptions 4.1–4.5 hold, then*

$$\boldsymbol{\zeta}\sqrt{N}\hat{\mathbf{A}} - N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\left(\hat{\mathbf{D}}_n - \mathbf{M}\right) = o_P(1).$$

Proof. Let

$$\mathbf{C} = \begin{pmatrix} \boldsymbol{\zeta}\mathbf{G}\boldsymbol{\zeta} & \mathbf{0}_{r\times p} & \mathbf{M} \\ \mathbf{0}_{p\times r} & \boldsymbol{\Lambda} & \mathbf{0}_{p\times 1} \\ \mathbf{M}^T & \mathbf{0}_{1\times p} & 1 \end{pmatrix}.$$

Using the fact that $\boldsymbol{\zeta}\mathbf{M} = \mathbf{M}$, one can verify via matrix multiplication that

$$\mathbf{C}^{-1} = \begin{pmatrix} \boldsymbol{\zeta}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta} & \mathbf{0}_{r\times p} & -\boldsymbol{\zeta}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M} \\ \mathbf{0}_{p\times r} & \boldsymbol{\Lambda}^{-1} & \mathbf{0}_{p\times 1} \\ -\mathbf{M}^T\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1} & \mathbf{0}_{1\times p} & 1 + \mathbf{M}^T\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\mathbf{M} \end{pmatrix}.$$

Since $N^{-1/2}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**}$ is bounded in probability, by (4.12) and Lemma 4.7 we have

$$\sqrt{N}\begin{pmatrix} \hat{\mathbf{A}} \\ \hat{\mathbf{B}} - \tilde{\mathbf{B}} \\ \hat{\mu} - \mu \end{pmatrix} - \mathbf{C}^{-1}N^{-1/2}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**} = o_P(1). \tag{4.29}$$

We observe that $\mathbf{C}^{-1}N^{-1/2}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**}$ can be expressed as

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\begin{pmatrix} \boldsymbol{\zeta}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta} & \mathbf{0}_{r\times p} & -\boldsymbol{\zeta}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M} \\ \mathbf{0}_{p\times r} & \boldsymbol{\Lambda}^{-1} & \mathbf{0}_{p\times 1} \\ -\mathbf{M}^T\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1} & \mathbf{0}_{1\times p} & 1 + \mathbf{M}^T\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\mathbf{M} \end{pmatrix}\begin{pmatrix} \hat{\mathbf{D}}_n \\ \hat{\mathbf{F}}_n \\ 1 \end{pmatrix}. \tag{4.30}$$

Notice that the first $r = p(p+1)/2$ elements of the vector in (4.30) are given by

$$\begin{aligned} &N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\boldsymbol{\zeta}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta} \quad \mathbf{0}_{r\times p} \quad -\boldsymbol{\zeta}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M}\right)\begin{pmatrix} \hat{\mathbf{D}}_n \\ \hat{\mathbf{F}}_n \\ 1 \end{pmatrix} \\ &= N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\boldsymbol{\zeta}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\hat{\mathbf{D}}_n - \boldsymbol{\zeta}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M}\right) \\ &= N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\boldsymbol{\zeta}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\left(\hat{\mathbf{D}}_n - \mathbf{M}\right). \end{aligned}$$

Therefore

$$\sqrt{N}\hat{\mathbf{A}} - N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\boldsymbol{\zeta}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\left(\hat{\mathbf{D}}_n - \mathbf{M}\right) = o_P(1). \tag{4.31}$$

The result is now obtained by multiplying (4.31) on the left by $\boldsymbol{\zeta}$. $\square$

Lemma 4.9. *If Assumptions 4.1–4.5 hold, then*

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\left(\hat{\mathbf{D}}_n - \mathbf{M}\right) \xrightarrow{d} N\left(0, \tau^2\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1}\right),$$

where

$$\tau^2 = \sigma^2 + \sum_{i=p+1}^{\infty} b_i^2\lambda_i$$

and $\sigma^2 = \mathrm{Var}\varepsilon_n$.

Proof. We prove this lemma in three steps. First we establish that

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\left(\boldsymbol{\zeta}\hat{\mathbf{D}}_n-\mathbf{M}\right)-\left(\mathbf{D}_n-\mathbf{M}\right)\right)=o_P\left(1\right). \tag{4.32}$$

In the second step we prove that

$$N^{-1/2}\sum_{n=1}^{N}\left(\mathbf{D}_n-\mathbf{M}\right)\left(\varepsilon_n^{**}-\varepsilon_n^{*}-\sum_{i=1}^{p}b_i\langle\bar{X}-\mu_X,\hat{c}_i\hat{v}_i\rangle\right)=o_P\left(1\right) \tag{4.33}$$

and

$$N^{-1/2}\sum_{n=1}^{N}\left(\mathbf{D}_n-\mathbf{M}\right)\langle\bar{X}-\mu_X,\hat{c}_i\hat{v}_i\rangle=o_P\left(1\right). \tag{4.34}$$

Combining (4.32), (4.33), and (4.34) we obtain immediately that

$$N^{-1/2}\sum_{n=1}^{N}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\left(\varepsilon_n^{**}\left(\boldsymbol{\zeta}\hat{\mathbf{D}}_n-\mathbf{M}\right)-\varepsilon_n^{*}\left(\mathbf{D}_n-\mathbf{M}\right)\right)=o_P\left(1\right).$$

Therefore, the lemma will be established by the third step:

$$N^{-1/2}\sum_{n=1}^{N}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\varepsilon_n^{*}\left(\mathbf{D}_n-\mathbf{M}\right)\xrightarrow{d}N\left(0,\tau^2\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\right). \tag{4.35}$$

We will now proceed to prove (4.32). The left side of (4.32) can be expressed elementwise as

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle X_n-\bar{X},\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_i\rangle\langle X_n^c,v_j\rangle\right)=o_P\left(1\right), \tag{4.36}$$

so it is sufficient to show that

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_i\rangle\langle X_n^c,v_j\rangle\right)=O_P\left(N^{-1/2}\right) \tag{4.37}$$

and

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle X_n-\bar{X},\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\langle X_n^c,\hat{c}_j\hat{v}_j\rangle\right)=o_P\left(1\right). \tag{4.38}$$

The left side of (4.37) is

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)+N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n^c,v_j\rangle\left(\langle X_n^c,\hat{c}_i\hat{v}_i\rangle-\langle X_n^c,v_i\rangle\right).$$

It follows from Assumptions 4.1–4.4 that both sets of random functions $\{\varepsilon_n X_n^c(t)X_n^c(s), 1\leq n\leq N\}$ and $\{X_n^c(u)X_n^c(t)X_n^c(s), 1\leq n\leq N\}$ are independent and identically distributed with zero mean so by the central limit theorem in Hilbert spaces we have

$$\left\|N^{-1/2}\sum_{n=1}^{N}\varepsilon_n X_n^c(t)X_n^c(s)\right\|_2=O_P\left(1\right)\quad\text{and}\quad\left\|N^{-1/2}\sum_{n=1}^{N}X_n^c(u)X_n^c(t)X_n^c(s)\right\|_2=O_P\left(1\right). \tag{4.39}$$

Next we write that

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)=\delta_1+\delta_2+\delta_3+\delta_4,$$

where, by (4.39), Lemma 4.1 and repeated applications of the Cauchy-Schwarz inequality, we have

$$\begin{aligned}|\delta_1|&=\left|N^{-1/2}\sum_{n=1}^{N}\varepsilon_n\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)\right|\\&\le\left|\left|N^{-1/2}\sum_{n=1}^{N}\varepsilon_n X_n^c(t)X_n^c(s)\hat{c}_i\hat{v}_i(t)\left(\hat{c}_j\hat{v}_j(s)-v_j(s)\right)\right|\right|_1\\&\le\left|\left|N^{-1/2}\sum_{n=1}^{N}\varepsilon_n X_n^c(t)X_n^c(s)\right|\right|_2||\hat{c}_j\hat{v}_j(s)-v_j(s)||_2\\&=O_P\left(N^{-1/2}\right),\end{aligned}$$

$$\begin{aligned}|\delta_2|&=\left|N^{-1/2}\sum_{n=1}^{N}\sum_{k=p+1}^{\infty}b_k\langle X_n^c,v_k\rangle\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)\right|\\&\le\left|\left|N^{-1/2}\sum_{n=1}^{N}X_n^c(u)X_n^c(t)X_n^c(s)\right|\right|_2\left|\left|\sum_{k=p+1}^{\infty}b_k v_k(u)\right|\right|_2||\hat{c}_j\hat{v}_j(s)-v_j(s)||_2\\&=O_P\left(N^{-1/2}\right),\end{aligned}$$

$$\begin{aligned}|\delta_3|&=\left|N^{-1/2}\sum_{n=1}^{N}\sum_{k=1}^{p}b_k\langle X_n^c,v_k-\hat{c}_k\hat{v}_k\rangle\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)\right|\\&\le\sum_{k=1}^{p}|b_k|\left|\left|N^{-1/2}\sum_{n=1}^{N}X_n^c(t)X_n^c(s)X_n^c(w)\right|\right|_2||v_k(w)-\hat{c}_k\hat{v}_k(w)||_2\,||\hat{c}_j\hat{v}_j(s)-v_j(s)||_2\\&=O_P\left(N^{-1}\right),\end{aligned}$$

and

$$\begin{aligned}|\delta_4|&=\left|N^{-1/2}\sum_{n=1}^{N}\sum_{k=1}^{p}b_k\langle\bar{X}-\mu_X,\hat{c}_k\hat{v}_k\rangle\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)\right|\\&\le\left|\sum_{k=1}^{p}b_k\langle\bar{X}-\mu_X,\hat{c}_k\hat{v}_k\rangle\right|\left|\left|N^{-1/2}\sum_{n=1}^{N}X_n^c(t)X_n^c(s)\right|\right|_2||\hat{c}_j\hat{v}_j(s)-v_j(s)||_2\\&=O_P\left(N^{-1/2}\right).\end{aligned}$$

Similarly,

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n^c,v_j\rangle\left(\langle X_n^c,\hat{c}_i\hat{v}_i\rangle-\langle X_n^c,v_i\rangle\right)=o_P\left(1\right),$$

and therefore (4.37) is proven.

We now establish (4.38). The left side of (4.38) is equal to

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle+N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n^c,\hat{c}_j\hat{v}_j\rangle\langle\mu_X-\bar{X},\hat{c}_i\hat{v}_i\rangle.$$

We write that

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle=\delta_5+\delta_6+\delta_7+\delta_8,$$

where, by the central limit theorem in Hilbert spaces, Lemma 4.1, and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}|\delta_5|&=\left|N^{-1/2}\sum_{n=1}^{N}\varepsilon_n\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\\&\le\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left\|N^{-1/2}\sum_{n=1}^{N}\varepsilon_n\left(X_n(s)-\bar{X}(s)\right)\right\|_2\\&=O_P\left(N^{-1/2}\right),\end{aligned}$$

$$\begin{aligned}|\delta_6|&=\left|N^{-1/2}\sum_{n=1}^{N}\sum_{k=p+1}^{\infty}b_k\langle X_n^c,v_k\rangle\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\\&\le\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|N^{-1/2}\sum_{n=1}^{N}\sum_{k=p+1}^{\infty}b_k\langle X_n^c,v_k\rangle\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\right|\\&=\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|N^{-1/2}\sum_{n=1}^{N}\int_0^1\int_0^1X_n^c(t)\left(X_n(s)-\bar{X}(s)\right)\hat{v}_i(s)\sum_{k=p+1}^{\infty}b_kv_k(t)\,ds\,dt\right|\\&=\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|N^{-1/2}\sum_{n=1}^{N}\int_0^1\int_0^1\left(X_n(t)-\bar{X}(t)\right)\left(X_n(s)-\bar{X}(s)\right)\hat{v}_i(s)\sum_{k=p+1}^{\infty}b_kv_k(t)\,ds\,dt\right|\\&=N^{1/2}\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|\int_0^1\int_0^1\hat{c}(t,s)\hat{v}_i(s)\sum_{k=p+1}^{\infty}b_kv_k(t)\,ds\,dt\right|\\&=N^{1/2}\hat{\lambda}_i\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|\int_0^1\hat{v}_i(t)\sum_{k=p+1}^{\infty}b_kv_k(t)\,dt\right|\\&=N^{1/2}\hat{\lambda}_i\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|\int_0^1\sum_{k=p+1}^{\infty}b_kv_k(t)\left(\hat{v}_i(t)-\hat{c}_iv_i(t)\right)\,dt\right|\\&\le N^{1/2}\hat{\lambda}_i\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left\|\sum_{k=p+1}^{\infty}b_kv_k(t)\right\|_2||\hat{v}_i(t)-\hat{c}_iv_i(t)||_2\\&=O_P\left(N^{-1/2}\right),\end{aligned}$$

$$
\begin{aligned}
|\delta_7| &= \left| N^{-1/2} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k \langle X_n^c, v_k - \hat{c}_k \hat{v}_k \rangle \langle X_n - \bar{X}, \hat{c}_i \hat{v}_i \rangle \langle \mu_X - \bar{X}, \hat{c}_j \hat{v}_j \rangle \right| \\
&\leq \left| \langle \mu_X - \bar{X}, \hat{c}_j \hat{v}_j \rangle \right| \left\| N^{-1/2} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k X_n^c(t) \left(X_n(s) - \bar{X}(s) \right) \right\|_2 ||v_k(t) - \hat{c}_k \hat{v}_k(t)||_2 \\
&= O_P\left(N^{-1/2}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
|\delta_8| &= \left| N^{-1/2} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k \langle \bar{X} - \mu_X, \hat{c}_k \hat{v}_k \rangle \langle X_n - \bar{X}, \hat{c}_i \hat{v}_i \rangle \langle \mu_X - \bar{X}, \hat{c}_j \hat{v}_j \rangle \right| \\
&\leq \left| \langle \mu_X - \bar{X}, \hat{c}_j \hat{v}_j \rangle \right| \left| \sum_{k=1}^{p} b_k \langle \bar{X} - \mu_X, \hat{c}_k \hat{v}_k \rangle \right| \left\| N^{-1/2} \sum_{n=1}^{N} \left(X_n(s) - \bar{X}(s) \right) \right\|_2 \\
&= O_P\left(N^{-1/2}\right).
\end{aligned}
$$

This proves (4.38), which also completes the proof of (4.36) and hence (4.32).

We proceed to the second step, which is the proof of (4.33) and (4.34). We express (4.33) elementwise as

$$
N^{-1/2} \sum_{n=1}^{N} \left(\langle X_n^c, v_i \rangle \langle X_n^c, v_j \rangle - \lambda_i 1\{i = j\} \right) \left(\sum_{k=1}^{p} b_k \langle X_n^c, v_k - \hat{c}_k \hat{v}_k \rangle \right) = o_P(1). \tag{4.40}
$$

We observe that by the central limit theorem in Hilbert spaces and Lemma 4.1 we have

$$
\begin{aligned}
\left| N^{-1/2} \sum_{n=1}^{N} \left(\sum_{k=1}^{p} b_k \langle X_n^c, v_k - \hat{c}_k \hat{v}_k \rangle \right) \right| &\leq \left\| N^{-1/2} \sum_{n=1}^{N} X_n^c(t) \right\|_2 \sum_{k=1}^{p} |b_k| \, ||v_k(t) - \hat{c}_k \hat{v}_k(t)||_2 \\
&= O_P\left(N^{-1/2}\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\left| N^{-1/2} \sum_{n=1}^{N} \langle X_n^c, v_i \rangle \langle X_n^c, v_j \rangle \left(\sum_{k=1}^{p} b_k \langle X_n^c, v_k - \hat{c}_k \hat{v}_k \rangle \right) \right| \\
&\qquad \leq \sum_{k=1}^{p} |b_k| \left\| N^{-1/2} \sum_{n=1}^{N} X_n^c(t) X_n^c(s) X_n^c(w) \right\|_2 ||v_k(w) - \hat{c}_k \hat{v}_k(w)||_2 \\
&= O_P\left(N^{-1/2}\right).
\end{aligned}
$$

This proves (4.40) and hence (4.33). Next, we establish (4.34). We can express (4.34) elementwise as

$$
N^{-1/2} \sum_{n=1}^{N} \left(\langle X_n^c, v_k \rangle \langle X_n^c, v_\ell \rangle - \lambda_k 1\{k = \ell\} \right) \langle \bar{X} - \mu_X, \hat{c}_i \hat{v}_i \rangle = o_P(1). \tag{4.41}
$$

Using the previous arguments, one can easily verify (4.41), establishing (4.34).

We will now finish the proof of the lemma by establishing (4.35) as the third step. Using Assumptions 4.1, 4.3, and (4.4), we see that ε_n^* has mean zero and variance given by

$$\begin{aligned} E\left(\varepsilon_n^*\right)^2 &= E\left(\varepsilon_1^2\right) + E\left(\sum_{i=p+1}^{\infty}\sum_{j=p+1}^{\infty} b_i b_j \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\right) \\ &= \sigma^2 + \sum_{i=p+1}^{\infty} b_i^2 E\left(\langle X_n^c, v_i\rangle^2\right) \\ &= \sigma^2 + \sum_{i=p+1}^{\infty} b_i^2 \lambda_i. \\ &= \tau^2 \end{aligned}$$

Therefore, $\varepsilon_n^*\left(\mathbf{D}_n - \mathbf{M}\right)$ is an iid sequence with mean zero and variance $\tau^2\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)$. The central limit theorem now proves (4.35), completing the proof of the lemma. $\square$

Lemma 4.10. *If Assumptions 4.2–4.5 are satisfied, then*

$$\begin{pmatrix} \hat{\mathbf{A}} \\ \hat{\mathbf{B}} \\ \hat{\mu} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \tilde{\mathbf{B}} \\ \mu \end{pmatrix} = O_P\left(N^{-1/2}\right). \tag{4.42}$$

In particular, we have

$$\|b_k v_k(t) - \hat{b}_k \hat{v}_k(t)\|_2 = O_P\left(N^{-1/2}\right) \tag{4.43}$$

and

$$\|\hat{a}_{i,j}\hat{v}_i(t)\hat{v}_j(s)\|_2 = O_P\left(N^{-1/2}\right), \tag{4.44}$$

where $\hat{a}_{i,j}$ and $\hat{b}_i$ are defined by

$$\hat{\mathbf{A}} = \operatorname{vech}\left(\{\hat{a}_{i,j}\left(2 - 1\{i=j\}\right), 1 \le i \le j \le p\}^T\right) \quad \textit{and} \quad \hat{\mathbf{B}} = \left(\hat{b}_1, \quad \hat{b}_2, \quad \ldots, \quad \hat{b}_p\right)^T.$$

Proof. Lemmas 4.8 and 4.9 imply that $\hat{\mathbf{A}} = O_P\left(N^{-1/2}\right)$. According to (4.29) and (4.30) we can prove that

$$\hat{\mathbf{B}} - \tilde{\mathbf{B}} = O_P\left(N^{-1/2}\right), \tag{4.45}$$

by showing that

$$\frac{1}{N}\sum_{n=1}^{N} \varepsilon_n^{**}\mathbf{\Lambda}^{-1}\hat{\mathbf{F}}_n = O_P\left(N^{-1/2}\right)$$

or equivalently that

$$\frac{1}{N}\sum_{n=1}^{N} \varepsilon_n^{**}\langle X_n - \bar{X}, \hat{v}_i\rangle = O_P\left(N^{-1/2}\right). \tag{4.46}$$

We note that

$$\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n-\bar{X},\hat{v}_i\rangle=\delta_9+\delta_{10}+\delta_{11}+\delta_{12},$$

where, following the arguments in the proof of Lemma 4.9, one can verify that

$$|\delta_9|=\left|\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n\langle X_n-\bar{X},\hat{v}_i\rangle\right|O_P\left(N^{-1/2}\right),$$

$$|\delta_{10}|=\left|\frac{1}{N}\sum_{n=1}^{N}\sum_{k=p+1}^{\infty}b_k\langle X_n^c,v_k\rangle\langle X_n-\bar{X},\hat{v}_i\rangle\right|=O_P\left(N^{-1/2}\right),$$

$$|\delta_{11}|=\left|\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{p}b_k\langle X_n^c,v_k-\hat{c}_k\hat{v}_k\rangle\langle X_n-\bar{X},\hat{v}_i\rangle\right|=O_P\left(N^{-1/2}\right),$$

and

$$|\delta_{12}|=\left|\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{p}b_k\langle \bar{X}-\mu_X,\hat{c}_k\hat{v}_k\rangle\langle X_n-\bar{X},\hat{v}_i\rangle\right|=O_P\left(N^{-1/2}\right).$$

This proves (4.46) and hence (4.45).

To complete the justification of (4.42), we need to show that

$$\hat{\mu}-\mu=O_P\left(N^{-1/2}\right). \tag{4.47}$$

Due to (4.29) and (4.30), (4.47) will be established by proving that

$$\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n^{**}\left(-\mathbf{M}^T\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\hat{\mathbf{D}}_n+1+\mathbf{M}^T\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\mathbf{M}\right)=O_P\left(N^{-1/2}\right). \tag{4.48}$$

To prove (4.48), it is sufficient to show

$$\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n^{**}\hat{\mathbf{D}}_n=O_P\left(N^{-1/2}\right) \tag{4.49}$$

and

$$\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n^{**}=O_P\left(N^{-1/2}\right). \tag{4.50}$$

Due to Lemma 4.9, (4.50) implies (4.49), so we prove only (4.50). We write that

$$\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n^{**}=\delta_{13}+\delta_{14}+\delta_{15}+\delta_{16},$$

where, by the central limit theorem in Hilbert spaces and Lemma 4.1, we have

$$|\delta_{13}|=\left|\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n\right|=O_P\left(N^{-1/2}\right),$$

$$|\delta_{14}| = \left| \frac{1}{N} \sum_{n=1}^{N} \sum_{k=p+1}^{\infty} b_k \langle X_n^c, v_k \rangle \right| \leq \left\| \frac{1}{N} \sum_{n=1}^{N} X_n^c(t) \right\|_2 \left\| \sum_{k=p+1}^{\infty} b_k v_k(t) \right\|_2 = O_P\left(N^{-1/2}\right),$$

$$|\delta_{15}| = \left| \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k \langle X_n^c, v_k - \hat{c}_k \hat{v}_k(t) \rangle \right| = O_P\left(N^{-1}\right),$$

and

$$|\delta_{16}| = \left| \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k \langle \bar{X} - \mu_X, \hat{c}_k \hat{v}_k \rangle \right| = O_P\left(N^{-1/2}\right).$$

This proves (4.50), which establishes (4.47) and completes the proof of (4.42).

Using (4.42) and Lemma 4.1, we will now show (4.43) and (4.44). We conclude from (4.42) that

$$\hat{b}_i - \hat{c}_i b_i = O_P\left(N^{-1/2}\right) \quad \text{and} \quad \hat{a}_{i,j} = O_P\left(N^{-1/2}\right).$$

Now, Lemma 4.1 yields that

$$\begin{aligned} \|b_k v_k(t) - \hat{b}_k \hat{v}_k(t)\|_2 &\leq \|b_k(v_k(t) - \hat{c}_k \hat{v}_k(t))\|_2 + \|(b_k \hat{c}_k - \hat{b}_k)\hat{v}_k(t)\|_2 \\ &\leq |b_k| \|v_k(t) - \hat{c}_k \hat{v}_k(t)\|_2 + |b_k \hat{c}_k - \hat{b}_k| \\ &= O_P\left(N^{-1/2}\right). \end{aligned}$$

Similarly,

$$\|\hat{a}_{i,j} \hat{v}_i(t) \hat{v}_j(s)\|_2 = O_P\left(N^{-1/2}\right).$$

This proves (4.43) and (4.44) and completes the proof of the lemma. □

Lemma 4.11. *If Assumptions 4.1–4.5 are satisfied, then*

$$\hat{\tau}^2 - \tau^2 = O_P\left(N^{-1/2}\right).$$

Proof. Since

$$\frac{1}{N} \sum_{n=1}^{N} \varepsilon_n^{*2} - \tau^2 \overset{a.s.}{\to} 0,$$

it is enough to show that

$$\frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n^2 - \varepsilon_n^{*2}\right) = O_P\left(N^{-1/2}\right). \tag{4.51}$$

Since

$$\frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n^2 - \varepsilon_n^{*2}\right) = \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n - \varepsilon_n^*\right)\left(\hat{\varepsilon}_n + \varepsilon_n^*\right) = \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n - \varepsilon_n^*\right)\hat{\varepsilon}_n + \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n - \varepsilon_n^*\right)\varepsilon_n^*,$$

(4.51) follows from

$$\left| \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n - \varepsilon_n^*\right) \varepsilon_n^* \right| = O_P\left(N^{-1/2}\right) \tag{4.52}$$

and

$$\left| \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n - \varepsilon_n^*\right) \hat{\varepsilon}_n \right| = O_P\left(N^{-1/2}\right). \tag{4.53}$$

We decompose (4.52) as

$$\frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n - \varepsilon_n^*\right) \varepsilon_n^* = \eta_1 + \eta_2 + \eta_3,$$

where

$$\begin{aligned}
\eta_1 &= \frac{1}{N} \sum_{n=1}^{N} \varepsilon_n^* \left(\mu - \hat{\mu}\right), \\
\eta_2 &= \frac{1}{N} \sum_{n=1}^{N} \varepsilon_n^* \sum_{i=1}^{p} \left(b_i \langle X_n^c, v_i \rangle - \hat{b}_i \langle X_n - \bar{X}, \hat{v}_i \rangle \right), \\
\eta_3 &= \frac{1}{N} \sum_{n=1}^{N} \varepsilon_n^* \sum_{i=1}^{p} \sum_{j=i}^{p} (2 - 1\{i = j\}) \left(a_{i,j} \langle X_n^c, v_i \rangle \langle X_n^c, v_j \rangle - \hat{a}_{i,j} \langle X_n - \bar{X}, \hat{v}_i \rangle \langle X_n - \bar{X}, \hat{v}_j \rangle \right).
\end{aligned}$$

It is clear that $\eta_1 = O_P\left(N^{-1}\right)$. We also see that $\eta_2 = \eta_{2,1} + \eta_{2,2} + \eta_{2,3} + \eta_{2,4}$, where

$$\begin{aligned}
\eta_{2,1} &= \frac{1}{N} \sum_{n=1}^{N} Y_n \sum_{i=1}^{p} \left(b_i \langle X_n^c, v_i \rangle - \hat{b}_i \langle X_n - \bar{X}, \hat{v}_i \rangle \right), \\
\eta_{2,2} &= -\frac{1}{N} \sum_{n=1}^{N} \mu \sum_{i=1}^{p} \left(b_i \langle X_n^c, v_i \rangle - \hat{b}_i \langle X_n - \bar{X}, \hat{v}_i \rangle \right), \\
\eta_{2,3} &= -\frac{1}{N} \sum_{n=1}^{N} \sum_{\ell=1}^{p} b_\ell \langle X_n^c, v_\ell \rangle \sum_{i=1}^{p} \left(b_i \langle X_n^c, v_i \rangle - \hat{b}_i \langle X_n - \bar{X}, \hat{v}_i \rangle \right), \\
\eta_{2,4} &= -\frac{1}{N} \sum_{n=1}^{N} \sum_{\ell=1}^{p} \sum_{k=\ell}^{p} (2 - 1\{k = \ell\}) a_{\ell,k} \langle X_n^c, v_\ell \rangle \langle X_n^c, v_k \rangle \sum_{i=1}^{p} \left(b_i \langle X_n^c, v_i \rangle - \hat{b}_i \langle X_n - \bar{X}, \hat{v}_i \rangle \right).
\end{aligned}$$

Applying (4.43) and the central limit theorem in Hilbert spaces we obtain that

$$
\begin{aligned}
|\eta_{2,1}| &= \left| \frac{1}{N} \sum_{n=1}^{N} Y_n \sum_{i=1}^{p} \Big(b_i \langle X_n^c, v_i \rangle - \hat{b}_i \langle X_n - \bar{X}, \hat{v}_i \rangle \Big) \right| \\
&\leq \sum_{i=1}^{p} \left|\left| \frac{1}{N} \sum_{n=1}^{N} Y_n \left(b_i X_n^c(t) v_i(t) - \hat{b}_i \left(X_n(t) - \bar{X}(t) \right) \hat{v}_i(t) \right) \right|\right|_1 \\
&\leq \sum_{i=1}^{p} \left|\left| \frac{1}{N} \sum_{n=1}^{N} Y_n X_n(t) \left(b_i v_i(t) - \hat{b}_i \hat{v}_i(t) \right) \right|\right|_1 \\
&\quad + \sum_{i=1}^{p} \left|\left| \frac{1}{N} \sum_{n=1}^{N} Y_n \left(b_i \mu_X(t) v_i(t) - \hat{b}_i \bar{X}(t) \hat{v}_i(t) \right) \right|\right|_1 \\
&\leq \sum_{i=1}^{p} \left|\left| \frac{1}{N} \sum_{n=1}^{N} Y_n X_n(t) \right|\right|_2 \left|\left| b_i v_i(t) - \hat{b}_i \hat{v}_i(t) \right|\right|_2 \\
&\quad + \sum_{i=1}^{p} \left|\left| \frac{1}{N} \sum_{n=1}^{N} Y_n \bar{X}(t) \left(\hat{b}_i \hat{v}_i(t) - b_i v_i(t) \right) \right|\right|_1 \\
&\quad + \sum_{i=1}^{p} \left|\left| \frac{1}{N} \sum_{n=1}^{N} Y_n b_i v_i(t) \left(\bar{X}(t) - \mu_X(t) \right) \right|\right|_1 \\
&\leq \sum_{i=1}^{p} \left|\left| \frac{1}{N} \sum_{n=1}^{N} Y_n X_n(t) \right|\right|_2 \left|\left| b_i v_i(t) - \hat{b}_i \hat{v}_i(t) \right|\right|_2 \\
&\quad + \sum_{i=1}^{p} \left|\left| \frac{1}{N} \sum_{n=1}^{N} Y_n \bar{X}(t) \right|\right|_2 \left|\left| \hat{b}_i \hat{v}_i(t) - b_i v_i(t) \right|\right|_2 \\
&\quad + \sum_{i=1}^{p} \left|\left| \frac{1}{N} \sum_{n=1}^{N} Y_n b_i v_i(t) \right|\right|_2 \left|\left| \bar{X}(t) - \mu(t) \right|\right|_2 \\
&= O_P\left(N^{-1/2} \right).
\end{aligned}
$$

In a like manner, one can verify that $\eta_{2,i} = O_P\left(N^{-1/2}\right), i = 2, 3, 4$. This proves that $\eta_2 = O_P\left(N^{-1/2}\right)$. In a similar fashion, one can show that $\eta_3 = O_P\left(N^{-1/2}\right)$. This proves (4.52). Following the previous arguments, one can establish (4.53), completing the proof of the lemma. □

4.7 Bibliography

Borggaard, C. and H. Thodberg (1992). Optimal Minimal Neural Interpretation of Spectra. *The Annals of Chemistry 64*, 545–551.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer.

Cai, T. and P. Hall (2006). Prediction in Functional Linear Regression. *The Annals of Statistics 34*, 2159–2179.

Cardot, H., F. Ferraty, A. Mas, and P. Sarda (2003). Testing Hypothesis in the Functional Linear Model. *Scandinavian Journal of Statistics 30*, 241–255.

Cardot, H. and P. Sarda (2011). Functional Linear Regression. In F. Ferraty and Y. Romain (Eds.), *The Oxford Handbook of Functional Data Analysis*, pp. 21–46. Oxford University Press.

Dauxois, J., A. Pousse, and Y. Romain (1982). Asymptotic Theory for Principal Component Analysis of a Vector Random Function. *Journal of Multivariate Analysis 12*, 136–154.

Fitzhugh, H. A. (1976). Analysis of Growth Curves and Strategies for Altering their Shapes. *Journal of Animal Science 33*, 1036–1051.

Frank, I. E. and J. H. Friedman (1993). A Statistical View of some Chemometrics Regression Tools. *Technometrics 35*, 109–135.

Hall, P. and J. Horowitz (2007). Methodology and Convergence Rates for Functional Linear Regression. *Annals of Statistics 35*, 70–91.

Kirkpatrick, M. and N. Heckman (1989). A Quantitiative Genetic Model for Growth, Shape, Reaction Norms and Other Infinite-dimensional Characters. *Journal of Mathematical Biology 27*, 429–450.

Mas, A. and B. Pumo (2011). Linear Processes for Functional Data. In F. Ferraty and Y. Romain (Eds.), *The Oxford Handbook of Functional Data Analysis*, pp. 47–71. Oxford University Press.

Müller, H.-G. and Y. Zhang (2005). Time-varying Functional Regression for Predicting Remaining Lifetime Distributions from Longitudinal Trajectories. *Biometrics 61*, 1064–1075.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. Springer.

Shen, Q. and J. Faraway (2004). An F Test for Linear Models with Functional Responses. *Statistica Sinica 14*, 1239–1257.

Wold, S. (1993). Discussion: PLS in Chemical Practice. *Technometrics 35*, 136–139.

Yao, F. and H.-G. Müller (2010). Functional Quadratic Regression. *Biometrika 97*, 49–64.

CHAPTER 5

DETECTING CHANGES IN FUNCTIONAL LINEAR MODELS[3]

We observe two sequences of curve which are connected via an integral operator. Our model includes linear models as well as autoregressive models in Hilbert spaces. We wish to test the null hypothesis that the operator did not change during the observation period. Our method is based on projecting the observations onto a suitably chosen finite dimensional space. The testing procedure is based on functionals of the weighted residuals of the projections. Since the quadratic form is based on estimating the long-term covariance matrix of the residuals, we also provide some results on Bartlett-type estimators.

5.1 Introduction

Suppose $\{X_n(t),\ n = 1, 2, \ldots, N\}$ and $\{Y_n(t),\ n = 1, 2, \ldots, N\}$ are sequences of random functions on $[0, 1]$ that satisfy the linear relationship

$$Y_n(t) = \int_0^1 \Psi_n(s, t) X_n(s)\, ds + \epsilon_n(t). \tag{5.1}$$

For example, $X_n(t)$ and $Y_n(t)$ may be the exchange rates of two currencies on day n at time t, where the trading day is normalized so that t ranges between 0 and 1. In other applications, X_n can be the temperature and Y_n the pollution level at a given location. If $\Psi_1 = \Psi_2 = \cdots = \Psi_N$, we say that the model is stable. However, as the underlying conditions change, the Ψ's may also change. Our estimates for the assumed common Ψ as well as our predictions and inferences based on the model would be flawed if we falsely assume that the Ψ's have not changed. To test the applicability of this model with an unchanging Ψ, we will test the null hypothesis,

$$H_0 : \Psi_1 = \Psi_2 = \cdots = \Psi_N, \tag{5.2}$$

---

[3]The content of this chapter is based on joint research with Lajos Horváth. It has been submitted to the Journal of Multivariate Analysis.

against the alternative

$$H_A: \quad \Psi_1 = \Psi_2 = \ldots = \Psi_{k_1^*} \neq \Psi_{k_1^*+1} = \ldots = \Psi_{k_r^*} \neq \Psi_{k_r^*+1} = \ldots = \Psi_N$$

with some unknown integers $k_1^*, \ldots k_r^*$. The k_i^*'s are called change-points, and the alternative, H_A, is that there are exactly r change-points. We assume that (5.1) and H_0 hold and that both $\{X_n\}$ and $\{Y_n\}$ are stationary sequences. The model with nonchanging (stable) Ψ has received considerable attention in the literature. If X_n and ϵ_n are independent sequences of independent processes, then (5.1) is a functional version of the classical linear model (Cardot et al., 2003; Chiou et al., 2004; Cai and Hall, 2006; Ferraty and Vieu, 2006). If $X_n = Y_{n-1}$, then we have the functional AR(1) model in (5.1) (Bosq, 2000; Kargin and Onatski, 2008; Horváth et al., 2010).

Let $C(s,t) = \operatorname{var}(X_n(t), X_n(s))$ and $D(s,t) = \operatorname{var}(Y_n(t), Y_n(s))$. Let $\{(v_j(s), \lambda_j),\ 1 \leq j \leq \infty\}$ and $\{(w_i(t), \tau_i),\ 1 \leq i \leq \infty\}$ be eigenfunction-eigenvalue pairs associated with $C(s,t)$ and $D(s,t)$ respectively. This means that $\tau_i w_i(t) = \int_0^1 D(t,s) w_i(s)\, ds$ and $\lambda_j v_j(s) = \int_0^1 C(s,t) v_j(t)\, dt$. Assume that λ_j is the j^{th} largest eigenvalue of $C(s,t)$ and that τ_i is the i^{th} largest eigenvalue of $D(s,t)$. It can be assumed that the eigenfunctions of $C(s,t)$ are orthonormal and also that the eigenfunctions of $D(s,t)$ are orthonormal. We assume that $\Psi \in L^2[0,1]^2$ and can therefore be expressed as

$$\Psi(s,t) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \psi_{i,j} v_j(s) w_i(t). \tag{5.3}$$

Using (5.3) we can write the model (5.1) as

$$\begin{aligned} Y_n(t) &= \int_0^1 \Psi_n(s,t) X_n(s)\, ds + \epsilon_n(t) \\ &= \int_0^1 \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \psi_{i,j} w_i(t) v_j(s) X_n(s)\, ds + \epsilon_n(t) \\ &= \sum_{i=1}^{q} \sum_{j=1}^{p} \psi_{i,j} w_i(t) \int_0^1 v_j(s) X_n(s)\, ds + \epsilon_n^*(t), \end{aligned} \tag{5.4}$$

where

$$\epsilon_n^*(t) = \epsilon_n(t) + \sum_{i=1}^{q} \sum_{j=p+1}^{\infty} \psi_{i,j} w_i(t) \int_0^1 v_j(s) X_n(s)\, ds + \sum_{i=q+1}^{\infty} \sum_{j=1}^{\infty} \psi_{i,j} w_i(t) \int_0^1 v_j(s) X_n(s)\, ds.$$

Equation (5.4) means that we keep the parts of Y_n and X_n which are explained by the first q and p principal components.

To reduce the dimensionality of the model we will project both sides of (5.4) onto the space spanned by the functions $\{w_i(t),\ 1 \leq i \leq q\}$. Doing this we obtain the linear model

$$\begin{pmatrix} \langle Y_n, w_1 \rangle \\ \langle Y_n, w_2 \rangle \\ \vdots \\ \langle Y_n, w_q \rangle \end{pmatrix} = \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & \dots & \psi_{1,p} \\ \psi_{2,1} & \psi_{2,2} & \dots & \psi_{2,p} \\ \vdots & \vdots & \dots & \vdots \\ \psi_{q,1} & \psi_{q,2} & \dots & \psi_{q,p} \end{pmatrix} \begin{pmatrix} \langle X_n, v_1 \rangle \\ \langle X_n, v_2 \rangle \\ \vdots \\ \langle X_n, v_p \rangle \end{pmatrix} + \begin{pmatrix} \langle \epsilon_n^*, w_1 \rangle \\ \langle \epsilon_n^*, w_2 \rangle \\ \vdots \\ \langle \epsilon_n^*, w_q \rangle \end{pmatrix}. \tag{5.5}$$

Instead of testing the null hypothesis, (5.2), exactly as it is stated, we would like to test if the coefficients $\{\psi_{i,j},\ 1 \leq i \leq q, 1 \leq j \leq p\}$ remained constant during the observation period. Essentially, we are testing the stability of $\Psi(s,t)$ over the space spanned by the most important principal components of the X_n's and the Y_n's. Equation (5.5) has the form of a linear model, but it is not a classical linear model because the regressors are random variables and are correlated with the errors. Unfortunately, we cannot use (5.5) directly, since the covariance functions, $D(s,t)$ and $C(s,t)$, and hence the eigenfunctions, $\{w_i(t),\ i = 1, 2, \dots q\}$ and $\{v_j(t),\ j = 1, 2, \dots, p\}$, are unknown. Instead, we will use the estimates $\hat{D}_N(s,t)$ and $\hat{C}_N(s,t)$ and their corresponding eigenfunctions, $\{\hat{w}_{i,N}(t),\ i = 1, 2, \dots q\}$ and $\{\hat{v}_{j,N}(s),\ j = 1, 2, \dots, p\}$, where

$$\hat{D}_N(s,t) = \frac{1}{N}\sum_{k=1}^{N}(Y_k(t) - \bar{Y}_N(t))(Y_k(s) - \bar{Y}_N(s)) \quad \text{with} \quad \bar{Y}_N(t) = \frac{1}{N}\sum_{i=1}^{N} Y_i(t),$$
$$\hat{C}_N(s,t) = \frac{1}{N}\sum_{k=1}^{N}(X_k(t) - \bar{X}_N(t))(X_k(s) - \bar{X}_N(s)) \quad \text{with} \quad \bar{X}_N(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t).$$

Eigenfunctions corresponding to unique eigenvalues are uniquely determined up to signs. For this reason, we cannot expect more than to have $\hat{w}_{i,N}$ be close to $\hat{d}_{i,N} w_i$ and $\hat{v}_{j,N}$ be close to $\hat{c}_{j,N} v_j$, where $\hat{d}_{i,N}$, $\hat{c}_{i,N}$ are random signs. In order to obtain a linear model similar to equation (5.5) that is useable, we must use our estimates for the eigenfunctions. We replace equation (5.4) with

$$Y_n(t) = \sum_{i=1}^{q}\sum_{j=1}^{p} \hat{d}_{i,N}\psi_{i,j}\hat{c}_{j,N}\hat{w}_{i,N}(t)\int_0^1 \hat{v}_{j,N}(s) X_n(s)\,ds + \epsilon_n^{**}(t), \tag{5.6}$$

where

$$\begin{aligned}\epsilon_n^{**}(t) =& \epsilon_n(t) + \sum_{i=1}^{q}\sum_{j=p+1}^{\infty} \psi_{i,j} w_i(t)\int_0^1 v_j(s)X_n(s)\,ds + \sum_{i=q+1}^{\infty}\sum_{j=1}^{\infty}\psi_{i,j}w_i(t)\int_0^1 v_j(s)X_n(s)\,ds \\ & - \sum_{i=1}^{q}\sum_{j=1}^{p}\hat{d}_{i,N}\psi_{i,j}\hat{c}_{j,N}\hat{w}_{i,N}(t)\int_0^1 \hat{v}_{j,N}(s)X_n(s)\,ds + \sum_{i=1}^{q}\sum_{j=1}^{p}\psi_{i,j}w_i(t)\int_0^1 v_j(s)X_n(s)\,ds.\end{aligned}$$

By projecting both sides of (5.6) onto the space spanned by the functions $\{\hat{w}_{j,N}(t),\ 1 \leq j \leq q\}$, we can replace the linear model (5.5) with the empirical linear model

$$\begin{pmatrix} \langle Y_n, \hat{w}_{1,N}\rangle \\ \langle Y_n, \hat{w}_{2,N}\rangle \\ \vdots \\ \langle Y_n, \hat{w}_{q,N}\rangle \end{pmatrix} = \begin{pmatrix} \hat{d}_{1,N}\psi_{1,1}\hat{c}_{1,N} & \hat{d}_1\psi_{1,2}\hat{c}_{2,N} & \dots & \hat{d}_{1,N}\psi_{1,p}\hat{c}_{p,N} \\ \hat{d}_{2,N}\psi_{2,1}\hat{c}_{1,N} & \hat{d}_{2,N}\psi_{2,2}\hat{c}_{2,N} & \dots & \hat{d}_{2,N}\psi_{2,p}\hat{c}_{p,N} \\ \vdots & \vdots & \dots & \vdots \\ \hat{d}_{q,N}\psi_{q,1}\hat{c}_{1,N} & \hat{d}_{q,N}\psi_{q,2}\hat{c}_{2,N} & \dots & \hat{d}_{q,N}\psi_{q,p}\hat{c}_{p,N} \end{pmatrix} \begin{pmatrix} \langle X_n, \hat{v}_{1,N}\rangle \\ \langle X_n, \hat{v}_{2,N}\rangle \\ \vdots \\ \langle X_n, \hat{v}_{p,N}\rangle \end{pmatrix} + \begin{pmatrix} \langle \epsilon_n^{**}, \hat{w}_{1,N}\rangle \\ \langle \epsilon_n^{**}, \hat{w}_{2,N}\rangle \\ \vdots \\ \langle \epsilon_n^{**}, \hat{w}_{q,N}\rangle \end{pmatrix}. \tag{5.7}$$

The signs $\{\hat{d}_{i,N},\ 1 \leq i \leq q\}$ and $\{\hat{c}_{j,N},\ 1 \leq j \leq p\}$ are computed from $X_1, X_2, \dots, X_N$ and $Y_1, Y_2, \dots, Y_N$ and they will not change during the testing procedure. Therefore, testing the stability of $\{\hat{d}_{i,N}\psi_{i,j}\hat{c}_{j,N},\ 1 \leq i \leq q, 1 \leq j \leq p\}$ is equivalent to testing the stability of $\{\psi_{i,j},\ 1 \leq i \leq q, 1 \leq j \leq p\}$.

Letting $\otimes$ be the Kronecker product, we can express equation (5.7) in a more condensed form:

$$\hat{\mathbf{Y}}^{(n)} = \hat{\mathbf{Z}}^{(n)}\boldsymbol{\beta} + \hat{\boldsymbol{\Delta}}^{(n)}, \qquad 1 \leq n \leq N, \tag{5.8}$$

where

$$\hat{\mathbf{Y}}^{(n)} = \begin{pmatrix} \langle Y_n, \hat{w}_{1,N}\rangle \\ \langle Y_n, \hat{w}_{2,N}\rangle \\ \vdots \\ \langle Y_n,\ \hat{w}_{q,N}\rangle \end{pmatrix}, \quad \hat{\boldsymbol{\Delta}}^{(n)} = \begin{pmatrix} \langle \epsilon_n^{**}, \hat{w}_{1,N}\rangle \\ \langle \epsilon_n^{**}, \hat{w}_{2,N}\rangle \\ \vdots \\ \langle \epsilon_n^{**}, \hat{w}_{q,N}\rangle \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \hat{d}_{1,N}\psi_{1,1}\hat{c}_{1,N} \\ \vdots \\ \hat{d}_{1,N}\psi_{1,p}\hat{c}_{p,N} \\ \hat{d}_{2,N}\psi_{2,1}\hat{c}_{1,N} \\ \vdots \\ \hat{d}_{q,N}\psi_{q,p}\hat{c}_{p,N} \end{pmatrix} = \mathrm{vec}(\{\hat{d}_{i,N}\psi_{i,j}\hat{c}_{j,N},\ 1 \leq i \leq q, 1 \leq j \leq p\}^T),$$

and

$$\hat{\mathbf{Z}}^{(n)} = \mathbf{I}_q \otimes \hat{\mathbf{M}}_n \quad \text{with} \quad \hat{\mathbf{M}}_n = (\langle X_n, \hat{v}_{1,N}\rangle, \dots, \langle X_n, \hat{v}_{p,N}\rangle).$$

The least squares estimator for $\boldsymbol{\beta}$ is defined by

$$\hat{\boldsymbol{\beta}}_N = \left(\hat{\mathbf{Z}}_N^T \hat{\mathbf{Z}}_N\right)^{-1} \hat{\mathbf{Z}}_N^T \hat{\mathbf{Y}}_N,$$

where the vectors $\hat{\mathbf{Y}}_{\lfloor Nt \rfloor}$ and the matrices $\hat{\mathbf{Z}}_{\lfloor Nt \rfloor}$ for each $t \in [0,1]$ are defined by

$$\hat{\mathbf{Y}}_{\lfloor Nt \rfloor} = \begin{pmatrix} \hat{\mathbf{Y}}^{(1)} \\ \hat{\mathbf{Y}}^{(2)} \\ \vdots \\ \hat{\mathbf{Y}}^{(\lfloor Nt \rfloor)} \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{Z}}_{\lfloor Nt \rfloor} = \begin{pmatrix} \hat{\mathbf{Z}}^{(1)} \\ \hat{\mathbf{Z}}^{(2)} \\ \vdots \\ \hat{\mathbf{Z}}^{(\lfloor Nt \rfloor)} \end{pmatrix}.$$

Our testing procedure is based on the cumulative sums process of the weighted residuals,

$$\tilde{\mathbf{V}}_N(t) = N^{-1/2} \left[\sum_{n=1}^{\lfloor Nt \rfloor} \left(\hat{\mathbf{Z}}^{(n)} \right)^T \tilde{\mathbf{Y}}^{(n)} - t \sum_{n=1}^{N} \left(\hat{\mathbf{Z}}^{(n)} \right)^T \tilde{\mathbf{Y}}^{(n)} \right], \quad t \in [0,1], \tag{5.9}$$

where $\tilde{\mathbf{Y}}^{(n)} = \hat{\mathbf{Y}}^{(n)} - \hat{\mathbf{Z}}^{(n)} \hat{\boldsymbol{\beta}}_N, 1 \leq n \leq N$ stands for the residuals.

5.2 Main Results

In this section we formally state all of the assumptions that we need and then we state our main theorem. Throughout this chapter we use $|\cdot|$ to mean the absolute value of a scalar or the largest of the absolute values of the elements of a vector or matrix. It will always be clear from the context which is meant.

Our first condition means that the processes X_n and ϵ_n are Bernoulli shifts:

Assumption 5.1. *$X_n(t)$ and $\epsilon_n(t)$ can be expressed as*

$$X_n(t) = a(\boldsymbol{\eta}_n(t), \boldsymbol{\eta}_{n-1}(t), \dots) \quad \text{and} \quad \epsilon_n(t) = b(\boldsymbol{\eta}_n(t), \boldsymbol{\eta}_{n-1}(t), \dots),$$

for some functionals a and b where $\{\boldsymbol{\eta}_k, -\infty < k < \infty\}$ are iid vector-valued random functions.

Assumption 5.1 implies immediately that the vector-valued process $(X_n, \epsilon_n), 1 \leq n < \infty$ is stationary and ergodic. If H_0 holds, then $(X_n, \epsilon_n, Y_n), 1 \leq n < \infty$ is also stationary and ergodic. We also require that the processes have at least 4 moments:

Assumption 5.2.

$$EX_n(t) = 0 \quad \text{and} \quad E\epsilon_n(t) = 0, \tag{5.10}$$

$$\int_0^1 EX_n^4(t)\, dt < \infty \quad \text{and} \quad \int_0^1 E\epsilon_n^4(t)\, dt < \infty. \tag{5.11}$$

Assumption 5.3. *$X_n(t)$ and $\epsilon_n(s)$ are uncorrelated, i.e. $EX_n(t)\epsilon_n(s) = 0$ for all $0 \leq t, s \leq 1$.*

Under assumption 5.1 one can even have long-range dependence among the observations. However, in this chapter we are interested only in weakly dependent sequences which is stated in the next assumption:

Assumption 5.4. *We assume that*

$$\sum_{1\le k<\infty}\left(E\int_0^1\left(X_n(t)-X_n^{(k)}(t)\right)^4dt\right)^{1/4}<\infty \tag{5.12}$$

and

$$\sum_{1\le k<\infty}\left(E\int_0^1\left(\epsilon_n(t)-\epsilon_n^{(k)}(t)\right)^4dt\right)^{1/4}<\infty \tag{5.13}$$

with

$$X_n^{(k)}(t)=a(\boldsymbol{\eta}_n(t),\boldsymbol{\eta}_{n-1}(t),\dots,\boldsymbol{\eta}_{n-k+1}(t),\boldsymbol{\eta}_{n,n-k}^{(k)}(t),\boldsymbol{\eta}_{n,n-k-1}^{(k)}(t),\dots)$$

and

$$\epsilon_n^{(k)}(t)=b(\boldsymbol{\eta}_n(t),\boldsymbol{\eta}_{n-1}(t),\dots,\boldsymbol{\eta}_{n-k+1}(t),\boldsymbol{\eta}_{n,n-k}^{(k)}(t),\boldsymbol{\eta}_{n,n-k-1}^{(k)}(t),\dots),$$

where $\{\boldsymbol{\eta}_{n,\ell}^{(k)},-\infty<k,\ell,n<\infty\}$ *are iid copies of* $\boldsymbol{\eta}_0$*.*

We note that, due to stationarity required by Assumption 5.1, it is enough to assume that (5.12) and (5.13) hold for at least one n. Hörmann and Kokoszka (2010) call the processes satisfying Assumption 5.4 L^2 m-decomposable processes. This property appeared first in Ibragimov (1962) and is used several times in Billingsley (1968). Aue et al. (2011) provide several examples when Assumptions 5.1 and 5.4 hold. For example, autoregressive, moving-average, linear processes in Hilbert spaces satisfy this condition. Also, the non-linear functional ARCH(1) model (Hörmann et al., 2010) and bilinear models (Hörmann and Kokoszka, 2010) also satisfy Assumption 5.4.

Our next assumption ensures that the p and q largest eigenvalues of C and D, respectively, are unique.

Assumption 5.5.

$$\lambda_1>\lambda_2>\cdots>\lambda_{p+1}$$

and

$$\tau_1>\tau_2>\cdots>\tau_{q+1}.$$

Assumption 5.6.

$$\int_0^1\int_0^1\Psi^4(s,t)\,dt\,ds<\infty.$$

We note that under Assumptions 5.2 and 5.6 we also have that $EY_n(t) = 0$ and $\int_0^1 EY_n^4(t)\,dt < \infty$. Let

$$\boldsymbol{\gamma}_\ell = \operatorname{vec}\left(\{\gamma_\ell(i,j), 1 \le i \le q, 1 \le j \le p\}^T\right),$$

where

$$\gamma_\ell(i,j) = \langle X_\ell, v_j\rangle\langle \epsilon_\ell, w_i\rangle + \langle X_\ell, v_j\rangle\langle X_\ell, u_i\rangle,$$

and

$$u_i(s) = \sum_{r=p+1}^{\infty} \psi_{i,r} v_r(s), \quad 1 \le i \le q.$$

Define $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = E\boldsymbol{\gamma}_0\boldsymbol{\gamma}_0^T + \sum_{l=1}^{\infty} E\boldsymbol{\gamma}_0\boldsymbol{\gamma}_\ell^T + \sum_{l=1}^{\infty} E\boldsymbol{\gamma}_\ell\boldsymbol{\gamma}_0^T.$$

We now define our detector as

$$V_N(t) = \tilde{\mathbf{V}}_N^T(t)\breve{\boldsymbol{\Sigma}}_N^{-1}\tilde{\mathbf{V}}_N(t),$$

where $\tilde{\mathbf{V}}_N(t)$ is defined in (5.9) and $\breve{\boldsymbol{\Sigma}}_N$ is an estimator (up to random signs) for $\boldsymbol{\Sigma}$. The Bartlett-type estimator that we propose for $\breve{\boldsymbol{\Sigma}}_N$ is a function of the estimators $\hat{v}_{j,N}(t)$ and $\hat{w}_{i,N}(t)$, which are estimators for $v(t)$ and $w(t)$ up to random signs. For this reason, we cannot expect that $\breve{\boldsymbol{\Sigma}}_N$ will be close to $\boldsymbol{\Sigma}$. The best we can expect is that $\boldsymbol{\zeta}_N\breve{\boldsymbol{\Sigma}}_N\boldsymbol{\zeta}_N$ will be close to $\boldsymbol{\Sigma}$, where $\boldsymbol{\zeta}_N$ is a matrix corresponding to the random signs, $\hat{c}_{j,N}$ and $\hat{d}_{i,N}$. This is described in assumption 5.7.

Next we introduce the diagonal matrices $\hat{\mathbf{C}}_N$ and $\hat{\mathbf{D}}_N$ which consists of the random signs, i.e. $\hat{\mathbf{C}}_N = \operatorname{diag}(\hat{c}_{1,N}, \ldots, \hat{c}_{p,N})$, $\hat{\mathbf{D}}_N = \operatorname{diag}(\hat{d}_{1,N}, \ldots, \hat{d}_{q,N})$ and $\boldsymbol{\zeta}_N = \hat{\mathbf{D}}_N \otimes \hat{\mathbf{C}}_N$.

Assumption 5.7. $\hat{\boldsymbol{\Sigma}}_N = \boldsymbol{\zeta}_N\breve{\boldsymbol{\Sigma}}_N\boldsymbol{\zeta}_N$ *is an estimator for* $\boldsymbol{\Sigma}$ *such that*

$$\left|\hat{\boldsymbol{\Sigma}}_N - \boldsymbol{\Sigma}\right| = o_P(1).$$

Note in particular that

$$\boldsymbol{\zeta}_N\boldsymbol{\gamma}_\ell = \operatorname{vec}\left(\{\hat{c}_{j,N}\hat{d}_{i,N}\gamma_\ell(i,j), 1 \le i \le q, 1 \le j \le p\}^T\right).$$

Note also that Assumption 5.7 and the continuous mapping theorem combined imply that $\hat{\boldsymbol{\Sigma}}_N^{-1} = \boldsymbol{\zeta}_N\breve{\boldsymbol{\Sigma}}_N^{-1}\boldsymbol{\zeta}_N \xrightarrow{P} \boldsymbol{\Sigma}^{-1}$.

Although any estimator satisfying Assumption 5.7 can be used, we recommend using a Bartlett-type estimator as $\breve{\boldsymbol{\Sigma}}_N$, which we will describe in Section 5.3.

Theorem 5.1. *If Assumptions 5.1-5.7 hold, then we have*

$$V_N(t) \xrightarrow{d} \sum_{\ell=1}^{pq} B_\ell^2(t),$$

where $\{B_\ell(t), \ell = 1, \ldots, pq\}$ *are iid standard Brownian bridges.*

The testing procedure can be based on Theorem 5.1, using functionals of $V_N(t)$. The distribution of functionals of the limit was considered by Kiefer (1959) who provided formulae for the distribution functions of the supremum and L_2 functionals of the limit. For tables, approximations and further discussion on the distribution of functionals of the limit we refer to Aue et al. (2009).

5.3 Bartlett-type estimators

In this section we discuss the estimation of the long-run covariance matrix of the sums of weakly dependent vectors. We start with estimators based on the sequence $\gamma_\ell, 1 \leq \ell \leq N$. Since $\mathbf{\Sigma}$ is the spectral density at 0, the kernel-type estimators discussed in Grenander and Rosenblatt (1957), Anderson (1971), Brillinger (1975), Priestley (1981), and Rosenblatt (1985) can be used. The estimator is defined by

$$\tilde{\mathbf{\Sigma}}_N = \sum_{k=-(N-1)}^{N-1} K(k/B_N)\boldsymbol{\phi}_{k,N},$$

where

$$\boldsymbol{\phi}_{k,N} = \frac{1}{N} \sum_{\ell=\max(1,1-k)}^{\min(N,N-k)} \boldsymbol{\gamma}_\ell \boldsymbol{\gamma}_{\ell+k}^T.$$

The kernel K satisfies the following condition:

Assumption 5.8.

- (i) $K(0) = 1$
- (ii) *K is a symmetric, Lipschitz function*
- (iii) *K has a bounded support*
- (iv) $\hat{K}$, *the Fourier transform of* K, *is also Lipschitz and integrable*

These conditions are mild, and they are satisfied by the most commonly used kernels, like the triangle of Bartlett and the polynomial kernel of Parzen (1961, 1967). Assumption 5.8(iii) makes the present proofs relatively technically simple and it could be replaced with the assumption that $K(x)$ decays sufficiently fast as $|x| \to \infty$. The next assumption is standard in the estimation of spectral densities and long term variances and covariances.

Assumption 5.9.

$$B_N \to \infty \quad and \quad B_N/N \to 0.$$

Jansson (2002) proved the consistency of covariance estimation for linear processes under the assumption $B_N = o(N^{1/2})$. Similarly, Hörmann and Kokoszka (2010) obtained consistency results for the estimation of the long run covariance matrices of the projections of functional observations assuming $B_N = o(N^{1/2})$. Liu and Wu (2010) established consistency results for estimation of spectral densities under Assumption 5.9.

Theorem 5.2. *If Assumptions 5.1-5.4, 5.6, 5.8 and 5.9 hold, then*

$$\tilde{\mathbf{\Sigma}}_N \xrightarrow{P} \mathbf{\Sigma}.$$

We would like to point out that the proof of Theorem 5.2 only requires that γ_ℓ is a Bernoulli shift with zero mean and finite second moment for which (5.28) holds.

The estimator, $\tilde{\mathbf{\Sigma}}_N$, cannot be computed since the variables $\boldsymbol{\gamma}_\ell$ are not observed directly and we need to estimate them from the sample. We have estimators for v_j as well as for w_i, but we will also need an estimator for ϵ_ℓ. We use the residuals to get inference on ϵ_ℓ:

$$\hat{\epsilon}_\ell(t) = Y_\ell(t) - \sum_{i=1}^{q}\sum_{j=1}^{p} \hat{\psi}_{i,j}\hat{w}_{i,N}(t)\langle X_\ell, \hat{v}_{j,N}\rangle,$$

where $\hat{\psi}_{i,j}$ is the $(i,j)^{th}$ element of $\hat{\boldsymbol{\beta}}_N$ when it is written in the matrix form, i.e. $\{\hat{\psi}_{i,j}, 1 \le i \le q, 1 \le j \le p\} = \mathrm{vec}^{-1}(\hat{\boldsymbol{\beta}}_N)$. Now $\boldsymbol{\gamma}_\ell$ will be replaced with

$$\hat{\boldsymbol{\gamma}}_\ell = \mathrm{vec}\left(\{\hat{\gamma}_\ell(i,j), 1 \le i \le q, 1 \le j \le p\}^T\right),$$

where

$$\hat{\gamma}_\ell(i,j) = \langle X_\ell, \hat{v}_{j,N}\rangle\langle \hat{\epsilon}_\ell, \hat{w}_{i,N}\rangle.$$

Now the Bartlett-type estimator is defined as

$$\breve{\mathbf{\Sigma}}_N = \sum_{k=-(N-1)}^{N-1} K(k/B_N)\hat{\boldsymbol{\phi}}_{k,N}, \tag{5.14}$$

where

$$\hat{\boldsymbol{\phi}}_{k,N} = \frac{1}{N}\sum_{\ell=\max(1,1-k)}^{\min(N,N-k)} \hat{\boldsymbol{\gamma}}_\ell\hat{\boldsymbol{\gamma}}_{\ell+k}^T.$$

The next result states that the proposed estimator satisfies Assumption 5.7.

Theorem 5.3. *If Assumptions 5.1–5.6, 5.8 hold and*

$$B_N \to \infty \quad and \quad B_N/N^{1/2} \to 0, \tag{5.15}$$

then Assumption 5.7 is satisfied.

5.4 Random Processes in Hilbert Spaces

In this section we summarize some basic results on random variables in Hilbert spaces which are used in the proofs. Let $\|\cdot\|$ denote the L^2-norm of functions defined on the unit interval, the unit square or the unit cube.

Theorem 5.4. *If Assumptions 5.1–5.4 hold, then we have*

$$\left\| \frac{1}{N^{1/2}} \sum_{n=1}^{N} X_n(t)\epsilon_n(s) \right\| = O_P(1), \tag{5.16}$$

$$\left\| \frac{1}{N^{1/2}} \sum_{n=1}^{N} (X_n(t)X_n(s) - C(t,s)) \right\| = O_P(1), \tag{5.17}$$

$$\left\| \frac{1}{N^{1/2}} \sum_{n=1}^{N} (\epsilon_n(t)\epsilon_n(s) - F(t,s))) \right\| = O_P(1), \tag{5.18}$$

with $F(t,s) = E(\epsilon_n(t)\epsilon_n(s))$. *If in addition Assumption 5.6 is also satisfied, then*

$$\left\| \frac{1}{N^{1/2}} \sum_{n=1}^{N} (Y_n(t)Y_n(s) - D(t,s)) \right\| = O_P(1). \tag{5.19}$$

Proof. It was pointed out in Hörmann and Kokoszka (2010) that the k-approximable property in Assumption 5.4 implies (5.17) and (5.18). Using (5.1), we get that the sums of $X_n(t)\epsilon_n(s)$ and $Y_n(t)Y_n(s)$ are also k-approximable so the rest of the result again follows from Theorem 3.1 of Hörmann and Kokoszka (2010). □

Theorem 5.5. *If Assumptions 5.1–5.6 hold, then we have*

$$\max_{1\le i\le q} \|\hat{w}_{i,N}(t) - \hat{d}_{i,N} w_i(t)\| = O_P\left(N^{-1/2}\right), \tag{5.20}$$

$$\max_{1\le j\le p} \|\hat{v}_{j,N}(t) - \hat{c}_{j,N} v_j(t)\| = O_P\left(N^{-1/2}\right) \tag{5.21}$$

and

$$\max_{1\le i\le q} |\hat{\tau}_{i,N} - \tau_i| = O_P\left(N^{-1/2}\right), \tag{5.22}$$

$$\max_{1\le j\le q} |\hat{\lambda}_{j,N} - \lambda_j| = O_P\left(N^{-1/2}\right). \tag{5.23}$$

Proof. Using Corollary 1.6 of Gohberg et al. (1990) we get that (5.20) follows from (5.19). According to Lemma 4.3 of Bosq (2000), (5.19) implies (5.22). Similarly, (5.17) yields (5.21) and (5.23). □

The next result is a uniform version of Theorem 5.4.

Theorem 5.6. *If Assumptions 5.1–5.4 and 5.6 hold, then we have*

$$\max_{1\le k\le N}\left\|\frac{1}{N^{1/2}}\sum_{n=1}^{k}X_n(t)\epsilon_n(s)\right\| = O_P(\log N), \tag{5.24}$$

$$\max_{1\le k\le N}\left\|\frac{1}{N^{1/2}}\sum_{n=1}^{k}(X_n(t)X_n(s)-C(t,s))\right\| = O_P(\log N), \tag{5.25}$$

$$\max_{1\le k\le N}\left\|\frac{1}{N^{1/2}}\sum_{n=1}^{k}(\epsilon_n(t)\epsilon_n(s)-F(t,s)))\right\| = O_P(\log N) \tag{5.26}$$

with $F(t,s) = E(\epsilon_n(t)\epsilon_n(s))$. If in addition Assumption 5.6 is also satisfied, then

$$\max_{1\le k\le N}\left\|\frac{1}{N^{1/2}}\sum_{n=1}^{k}(Y_n(t)Y_n(s)-D(t,s))\right\| = O_P(\log N). \tag{5.27}$$

Proof. Following the proof in Section A.1 in Hörmann and Kokoszka (2010) one can easily verify that there is an integrable function $g(t,s)$ such that

$$E\left(\sum_{n=1}^{k}X_n(t)\epsilon_n(s)\right)^2 \le kg(t,s).$$

Hence by Menshov's inequality (Móricz, 1976) we have that

$$E\max_{1\le k\le N}\left(\sum_{n=1}^{k}X_n(t)\epsilon_n(s)\right)^2 \le (\log N)^2 Ng(t,s),$$

implying (5.24). Similar arguments yield (5.25)-(5.27). □

The next results establish the weak convergence of the sum of the γ_ℓ's.

Theorem 5.7. *If Assumptions 5.1–5.4 and 5.6 hold, then*

$$\frac{1}{N^{1/2}}\sum_{\ell=1}^{\lfloor Nt\rfloor}\boldsymbol{\gamma}_\ell \quad \overset{\mathcal{D}^{pq}[0,1]}{\longrightarrow} \quad \mathbf{W}_{\boldsymbol{\Sigma}}(t),$$

where $\mathbf{W}_{\boldsymbol{\Sigma}}$ is a pq dimensional Brownian motion with zero mean and $E(\mathbf{W}_{\boldsymbol{\Sigma}}(t)\mathbf{W}_{\boldsymbol{\Sigma}}(s)^T) = \min(t,s)\boldsymbol{\Sigma}$.

Proof. First we note that Assumptions 5.1–5.4 imply that

$$\sum_{m=1}^{\infty}\left(E(\gamma_\ell(i)-\gamma_\ell^{(m)}(i))^2\right)^{1/2}<\infty, \tag{5.28}$$

where $\gamma_\ell(i)$ and $\gamma_\ell^{(m)}(i)$ are the i^{th} coordinates of the vectors $\boldsymbol{\gamma}_\ell$ and $\boldsymbol{\gamma}_\ell^{(m)}$ with

$$\boldsymbol{\gamma}_\ell^{(m)}=\mathrm{vec}(\{\gamma_\ell^{(m)}(i,j),1\le i\le q,1\le j\le p\}^T),$$

and

$$\gamma_\ell^{(m)}(i,j)=\langle X_\ell^{(m)},v_j\rangle\langle\epsilon_\ell^{(m)},w_i\rangle+\langle X_\ell^{(m)},v_j\rangle\langle X_\ell^{(m)},u_i\rangle.$$

The result now follows immediately from Theorem A.1 of Aue et al. (2009). □

5.5 Proof of Theorem 5.1

First we outline the proof of Theorem 5.1. Using the definition of the residual vectors we can write that

$$\begin{aligned}\tilde{\mathbf{V}}_N(t)&=N^{-1/2}\left(\left(\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Y}}_{\lfloor Nt\rfloor}-\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}\hat{\boldsymbol{\beta}}_N\right)-t\left(\hat{\mathbf{Z}}_N^T\hat{\mathbf{Y}}_N-\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N\hat{\boldsymbol{\beta}}_N\right)\right)\\&=N^{-1/2}\left(\left(\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\boldsymbol{\Delta}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\boldsymbol{\Delta}}_N\right)+\left(\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N\right)\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_N\right)\right)\\&=N^{-1/2}\left(\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\boldsymbol{\Delta}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\boldsymbol{\Delta}}_N\right)+\left(\frac{\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N}{N}\right)\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_N\right)\sqrt{N},\end{aligned} \tag{5.29}$$

with

$$\hat{\boldsymbol{\Delta}}_{\lfloor Nt\rfloor}=\begin{pmatrix}\hat{\boldsymbol{\Delta}}^{(1)}\\\hat{\boldsymbol{\Delta}}^{(2)}\\\vdots\\\hat{\boldsymbol{\Delta}}^{(\lfloor Nt\rfloor)}\end{pmatrix}.$$

We show that

$$\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_N\right)\sqrt{N}=O_P(1), \tag{5.30}$$

(cf. Lemma 5.10) and we prove in Lemma 5.2 that

$$\sup_{t\in[0,1]}\left|\frac{\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N}{N}\right|=o_P(1). \tag{5.31}$$

Combining (5.30) and (5.31) we conclude that

$$\sup_{t\in[0,1]}\left|\left(\frac{\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N}{N}\right)\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_N\right)\sqrt{N}\right|=o_P(1).$$

Thus we see that $N^{-1/2}\left(\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\boldsymbol{\Delta}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\boldsymbol{\Delta}}_N\right)$ is the leading term while the remainder can be disregarded when considering the limiting distribution of our cumulative sum process

(5.9).

We now start with the proof of (5.31).

Lemma 5.1. *If Assumptions 5.1–5.5 hold, then we have*

$$\frac{1}{k}\sum_{n=1}^{k}\langle X_n, v_i\rangle\langle X_n, v_j\rangle \stackrel{a.s.}{\to} \lambda_i\, 1\{i=j\} \quad as \quad k\to\infty.$$

Proof. We recall that $X_n(t)$ is stationary and ergodic. Thus the ergodic theorem shows us that as $k\to\infty$

$$\begin{aligned}\frac{1}{k}\sum_{n=1}^{k}\langle X_n, v_i\rangle\langle X_n,\ v_j\rangle &\stackrel{a.s.}{\to} E\int_0^1 X_n(s)v_i(s)\,ds\int_0^1 X_n(t)v_j(t)\,dt\\ &= E\int_0^1 v_j(t)\int_0^1 v_i(s)X_n(t)X_n(s)\,ds\,dt\\ &= \int_0^1 v_j(t)\int_0^1 v_i(s)E\left(X_n(t)X_n(s)\right)\,ds\,dt\\ &= \int_0^1 v_j(t)\int_0^1 v_i(s)C(s,t)\,ds\,dt\\ &= \int_0^1 v_j(t)\lambda_i v_i(t)\,dt\\ &= \lambda_i\, 1\{i=j\},\end{aligned}$$

completing the proof.

□

Lemma 5.2. *If Assumptions 5.1–5.5 hold, then we have*

$$\frac{1}{N}\sup_{t\in[0,1]}\left|\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor} - t\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N\right| = o_P\left(1\right) \tag{5.32}$$

and

$$\frac{1}{N}\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N \stackrel{P}{\to} \mathbf{C} = \mathbf{I}_q\otimes\mathbf{\Lambda}, \tag{5.33}$$

where $\mathbf{\Lambda} = diag(\lambda_1,\lambda_2,\ldots,\lambda_p)$.

Proof. First we show that for $\delta>0$ and $\gamma>0$ there are K_0 and N_0 such that

$$P\left(\sup_{K_0\le k\le N}\left|\frac{1}{k}\sum_{n=1}^{k}\langle X_n,\hat{v}_{i,N}\rangle\langle X_n,\hat{v}_{j,N}\rangle - \lambda_i 1\{i=j\}\right| > \delta\right)\le\gamma, \tag{5.34}$$

if $N\ge N_0$. Note that by the Cauchy-Schwarz inequality we have

$$\left|\frac{1}{k}\sum_{n=1}^{k}\big(\langle X_n,\hat{v}_{i,N}\rangle\langle X_n,\hat{v}_{j,N}\rangle - \langle X_n,\hat{c}_{i,N}v_i\rangle\langle X_n,\hat{c}_{j,N}v_j\rangle\big)\right|$$

$$\leq \frac{1}{k}\sum_{n=1}^{k}\|X_n\|^2(\|\hat{v}_{i,N}-\hat{c}_{i,N}v_i\|+\|\hat{v}_{j,N}-\hat{c}_{j,N}v_j\|).$$

Using the ergodic theorem we get that

$$\sup_{1\leq k<\infty}\frac{1}{k}\sum_{n=1}^{k}\|X_n\|^2<\infty \quad \text{a.s.},$$

so (5.34) follows from Theorem 5.5 and Lemma 5.1.

Assume $N>N_0$. It now follows that

$$\begin{aligned}
&P\left(\sup_{t\in[0,1]}\left|\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N\right|>4N\delta\right)\\
&\leq P\left(\sup_{0\leq t\leq K_0/N}\left|\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N\right|+\sup_{K_0/N\leq t\leq 1}\left|\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N\right|>4N\delta\right)\\
&\leq P\left(\sup_{0\leq t\leq K_0/N}\left|\frac{\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N}{N}\right|+\sup_{K_0/N\leq t\leq 1}\left|\frac{\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}}{Nt}-\frac{\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N}{N}\right|>4\delta\right)\\
&\leq P\left(\sup_{0\leq t\leq K_0/N}\left|\frac{\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}-t\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N}{N}\right|+\sup_{K_0/N\leq t\leq 1}\left|\frac{\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}}{Nt}-\mathbf{C}\right|+\left|\frac{\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N}{N}-\mathbf{C}\right|>4\delta\right)\\
&\leq P\left(\max_{1\leq k\leq K_0}\left|\frac{\hat{\mathbf{Z}}_k^T\hat{\mathbf{Z}}_k}{N}\right|+\left|\frac{K_0\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N}{N^2}\right|+\max_{K_0\leq k\leq N}\left|\frac{\hat{\mathbf{Z}}_k^T\hat{\mathbf{Z}}_k}{k}-\mathbf{C}\right|+\left|\frac{\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N}{N}-\mathbf{C}\right|>4\delta\right).
\end{aligned}$$

For every K_0 we have that $P(\max_{1\leq k\leq K_0}|\hat{\mathbf{Z}}_k^T\hat{\mathbf{Z}}_k|/N>\delta)\to 0$ and by (5.34) $P(|K_0\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N|/N^2>\delta)\to 0$ as $N\to\infty$. Using (5.34) again we conclude $P(\max_{K_0\leq k\leq N}|\hat{\mathbf{Z}}_k^T\hat{\mathbf{Z}}_k/k-\mathbf{C}|>\delta)\leq\gamma$. Since γ and δ can be chosen as small as we wish, Lemma 5.2 is established.

□

We continue with the properties of $\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{\Delta}}_{\lfloor Nt\rfloor}$. First we observe that

$$\begin{aligned}
\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{\Delta}}_{\lfloor Nt\rfloor}&=\sum_{\ell=1}^{\lfloor Nt\rfloor}\left(\hat{\mathbf{Z}}^{(\ell)}\right)^T\hat{\mathbf{\Delta}}^{(\ell)}\\
&=\sum_{\ell=1}^{\lfloor Nt\rfloor}\operatorname{vec}\left(\{\langle X_\ell,\hat{v}_{j,N}\rangle\langle\epsilon_\ell^{**},\ \hat{w}_{i,N}\rangle,1\leq i\leq q,1\leq j\leq p\}^T\right).
\end{aligned}\tag{5.35}$$

We note that

$$\epsilon_\ell^{**}(t)=\epsilon_\ell(t)+\eta_{\ell,1}(t)+\eta_{\ell,2}(t)+\eta_{\ell,3}(t)+\eta_{\ell,4}(t)+\eta_{\ell,5}(t),$$

with

$$\begin{aligned}
\eta_{n,1}(t) &= \sum_{i=q+1}^{\infty}\sum_{j=1}^{\infty}\psi_{i,j}w_i(t)\int_0^1 v_j(s)X_n(s)\,ds = \sum_{i=q+1}^{\infty}\sum_{j=1}^{\infty}\psi_{i,j}w_i(t)\langle v_j, X_n\rangle,\\
\eta_{n,2}(t) &= \sum_{i=1}^{q}\sum_{j=p+1}^{\infty}\psi_{i,j}w_i(t)\int_0^1 v_j(s)X_n(s)\,ds = \sum_{i=1}^{q}\sum_{j=p+1}^{\infty}\psi_{i,j}w_i(t)\langle v_j, X_n\rangle,\\
\eta_{n,3}(t) &= \sum_{i=1}^{q}\sum_{j=1}^{p}\hat{d}_{i,N}\psi_{i,j}\hat{c}_{j,N}\hat{d}_{i,N}w_i(t)\int_0^1 \left(\hat{c}_{j,N}v_j(s) - \hat{v}_{j,N}(s)\right)X_n(s)\,ds\\
&= \sum_{i=1}^{q}\sum_{j=1}^{p}\hat{d}_{i,N}\psi_{i,j}\hat{c}_{j,N}\hat{d}_{i,N}w_i(t)\langle\left(\hat{c}_{j,N}v_j - \hat{v}_{j,N}\right),\ X_n\rangle,\\
\eta_{n,4}(t) &= \sum_{i=1}^{q}\sum_{j=1}^{p}\hat{d}_{i,N}\psi_{i,j}\hat{c}_{j,N}\left(\hat{d}_{i,N}w_i(t) - \hat{w}_{i,N}(t)\right)\int_0^1 \hat{c}_{j,N}v_j(s)X_n(s)\,ds\\
&= \sum_{i=1}^{q}\sum_{j=1}^{p}\hat{d}_{i,N}\psi_{i,j}\hat{c}_{j,N}\left(\hat{d}_{i,N}w_i(t) - \hat{w}_{i,N}(t)\right)\langle\hat{c}_{j,N}v_j, X_n\rangle,\\
\eta_{n,5}(t) &= \sum_{i=1}^{q}\sum_{j=1}^{p}\hat{d}_{i,N}\psi_{i,j}\hat{c}_{j,N}\left(\hat{w}_{i,N}(t) - \hat{d}_{i,N}w_i(t)\right)\int_0^1 \left(\hat{c}_{j,N}v_j(s) - \hat{v}_{j,N}(s)\right)X_n(s)\,ds\\
&= \sum_{i=1}^{q}\sum_{j=1}^{p}\hat{d}_{i,N}\psi_{i,j}\hat{c}_{j,N}\left(\hat{w}_{i,N}(t) - \hat{d}_{i,N}w_i(t)\right)\langle\left(\hat{c}_{j,N}v_j - \hat{v}_{j,N}\right), X_n\rangle.
\end{aligned}$$

In particular, we can write

$$\begin{aligned}
\langle\epsilon_\ell^{**},\ \hat{w}_{i,N}\rangle &= \langle\epsilon_\ell, \hat{w}_{i,N}\rangle + \langle\eta_{\ell,1}, \hat{w}_{i,N}\rangle + \langle\eta_{\ell,2}, \hat{w}_{i,N}\rangle\\
&\quad + \langle\eta_{\ell,3}, \hat{w}_{i,N}\rangle + \langle\eta_{\ell,4}, \hat{w}_{i,N}\rangle + \langle\eta_{\ell,5}, \hat{w}_{i,N}\rangle.
\end{aligned} \tag{5.36}$$

We show that $\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\mathbf{\Delta}}_{\lfloor Nt\rfloor}$ can be written as the sum of weakly dependent variables and an additional term which is just t times a random variable matrix. The additional term reflects the replacement of Ψ with a finite sum and the estimation of the eigenfunctions $\{w_i, 1 \le i \le q\}$ and $\{v_j, 1 \le i \le p\}$. The drift term is given by

$$\mathbf{R}_N = \mathrm{vec}\left(\{R_N(i,j), 1 \le i \le q, 1 \le j \le p\}^T\right),$$

where

$$
\begin{aligned}
R_N(i,j) &= R_N^{(1)}(i,j) + R_N^{(2)}(i,j) + R_N^{(3)}(i,j) + R_N^{(4)}(i,j),\\
R_N^{(1)}(i,j) &= \hat{c}_{j,N}\lambda_j \sum_{r=q+1}^{\infty} \psi_{r,j} \int_0^1 w_r(x)\left(\hat{w}_{i,N}(x) - \hat{d}_{i,N} w_i(x)\right) dx,\\
R_N^{(2)}(i,j) &= \hat{d}_{i,N} \int_0^1 \left(\hat{v}_{j,N}(z) - \hat{c}_{j,N} v_j(z)\right) \sum_{n=p+1}^{\infty} \psi_{i,n}\lambda_n v_n(z)\, dz,\\
R_N^{(3)}(i,j) &= \hat{c}_{j,N}\hat{d}_{i,N}\lambda_j \sum_{n=1}^{p} \psi_{i,n}\hat{c}_{n,N} \int_0^1 \left(\hat{c}_{n,N} v_n(s) - \hat{v}_{n,N}(s)\right) v_j(s)\, ds,\\
R_N^{(4)}(i,j) &= \hat{c}_{j,N}\hat{d}_{i,N}\lambda_j \sum_{r=1}^{q} \hat{d}_{r,N}\psi_{r,j} \int_0^1 w_i(x)\left(\hat{d}_{r,N} w_r(x) - \hat{w}_{r,N}(x)\right) dx.
\end{aligned}
$$

Lemma 5.3. *If Assumptions 5.1–5.5 hold, then we have*

$$
\sup_{t\in[0,1]} \left| \sum_{\ell=1}^{\lfloor Nt \rfloor} \langle X_\ell, \hat{v}_{j,N}\rangle\langle \epsilon_\ell,\ \hat{w}_{i,N}\rangle - \hat{c}_{j,N}\hat{d}_{i,N} T^{(1)}_{\lfloor Nt \rfloor}(i,j) \right| = O_P(\log N),
$$

where

$$
T^{(1)}_{\lfloor Nt \rfloor}(i,j) = \sum_{\ell=1}^{\lfloor Nt \rfloor} \langle X_\ell, v_j\rangle\langle \epsilon_\ell,\ w_i\rangle.
$$

Proof. We note that

$$
\begin{aligned}
&\sup_{t\in[0,1]} \left| \sum_{\ell=1}^{\lfloor Nt \rfloor} \langle X_\ell, \hat{v}_{j,N}\rangle\langle \epsilon_\ell, \hat{w}_{i,N}\rangle - \sum_{\ell=1}^{\lfloor Nt \rfloor} \langle X_\ell, \hat{c}_{j,N} v_j\rangle\langle \epsilon_\ell, \hat{d}_{i,N} w_i\rangle \right|\\
&\le \sup_{t\in[0,1]} \left| \sum_{\ell=1}^{\lfloor Nt \rfloor} \langle X_\ell, \hat{v}_{j,N} - \hat{c}_{j,N} v_j\rangle\langle \epsilon_\ell, \hat{w}_{i,N}\rangle \right| + \sup_{t\in[0,1]} \left| \sum_{\ell=1}^{\lfloor Nt \rfloor} \langle X_\ell, \hat{c}_{j,N} v_j\rangle\langle \epsilon_\ell, \hat{w}_{i,N} - \hat{d}_{i,N} w_i\rangle \right|.
\end{aligned}
$$

Using the Cauchy-Schwarz inequality we get that

$$
\begin{aligned}
\sup_{t\in[0,1]} &\left| \sum_{\ell=1}^{\lfloor Nt \rfloor} \langle X_\ell, \hat{v}_{j,N} - \hat{c}_{j,N} v_j\rangle\langle \epsilon_\ell, \hat{w}_{i,N}\rangle \right|\\
&\le \sup_{t\in[0,1]} \left\| \sum_{\ell=1}^{\lfloor Nt \rfloor} X_\ell(x)\epsilon_\ell(s) \right\| \left(\|\hat{v}_{j,N} - \hat{c}_{j,N} v_j\| \|\hat{w}_{i,N}\|\right)\\
&= O_P(\log N),
\end{aligned}
$$

on account of (5.21), Theorem 5.6 and $\|\hat{w}_{i,N}\| = 1$. Similar arguments give that

$$
\sup_{t\in[0,1]} \left| \sum_{\ell=1}^{\lfloor Nt \rfloor} \langle X_\ell, \hat{c}_{j,N} v_j\rangle\langle \epsilon_\ell, \hat{w}_{i,N} - \hat{d}_{i,N} w_i\rangle \right| = O_P(\log N),
$$

completing the proof of the lemma. □

Lemma 5.4. *If Assumptions 5.1–5.6 hold, then we have*

$$\sup_{t\in[0,1]}\left|\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,\hat v_{j,N}\rangle\langle\eta_{\ell,1},\hat w_{i,N}\rangle-\lfloor Nt\rfloor R_N^{(1)}(i,j)\right|=O_P\left(\log N\right).$$

Proof. Using the orthogonality of the w_i's we get that

$$\begin{aligned}\langle\eta_{\ell,1},w_i\rangle&=\int_0^1\sum_{r=q+1}^{\infty}\sum_{n=1}^{\infty}\psi_{r,n}w_r(x)\left\{\int_0^1 v_n(s)X_\ell(s)\,ds\right\}w_i(x)\,dx\\&=\sum_{r=q+1}^{\infty}\sum_{n=1}^{\infty}\psi_{r,n}\int_0^1 v_n(s)X_\ell(s)\,ds\left(\int_0^1 w_i(x)w_r(x)\,dx\right)\\&=0.\end{aligned}$$

Therefore we have

$$\begin{aligned}&\langle X_\ell,\hat v_{j,N}\rangle\langle\eta_{\ell,1},\hat w_{i,N}\rangle\\&\quad=\langle X_\ell,\hat v_{j,N}-\hat c_{j,N}v_j\rangle\langle\eta_{\ell,1},\hat w_{i,N}\rangle+\langle X_\ell,\hat c_{j,N}v_j\rangle\langle\eta_{\ell,1},\hat w_{i,N}\rangle\\&\quad=\langle X_\ell,\hat v_{j,N}-\hat c_{j,N}v_j\rangle\langle\eta_{\ell,1},\hat w_{i,N}-\hat d_{i,N}w_i\rangle+\langle X_\ell,\hat c_{j,N}v_j\rangle\langle\eta_{\ell,1},\hat w_{i,N}-\hat d_{i,N}w_i\rangle.\end{aligned}$$

Now,

$$\begin{aligned}&\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,\hat c_{j,N}v_j\rangle\langle\eta_{\ell,1},\hat w_{i,N}-\hat d_{i,N}w_i\rangle\\&=\hat c_{j,N}\sum_{\ell=1}^{\lfloor Nt\rfloor}\int_0^1\int_0^1 X_\ell(z)v_j(z)\sum_{r=q+1}^{\infty}\sum_{n=1}^{\infty}\psi_{r,n}w_r(x)\\&\qquad\times\int_0^1 v_n(s)X_\ell(s)\left(\hat w_{i,N}(x)-\hat d_{i,N}w_i(x)\right)\,dz\,ds\,dx\\&=\hat c_{j,N}\int_0^1\int_0^1\int_0^1 v_j(z)\sum_{r=q+1}^{\infty}\sum_{n=1}^{\infty}\psi_{r,n}w_r(x)v_n(s)\left(\hat w_{i,N}(x)-\hat d_{i,N}w_i(x)\right)\\&\qquad\times\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(z)X_\ell(s)\,dz\,ds\,dx\\&=A_{\lfloor Nt\rfloor}^{(1)}+A_{\lfloor Nt\rfloor}^{(2)},\end{aligned}$$

where

$$\begin{aligned}A_{\lfloor Nt\rfloor}^{(1)}=\hat c_{j,N}&\int_0^1\int_0^1\int_0^1 v_j(z)\sum_{r=q+1}^{\infty}\sum_{n=1}^{\infty}\psi_{r,n}w_r(x)v_n(s)\left(\hat w_{i,N}(x)-\hat d_{i,N}w_i(x)\right)\\&\times\left(\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(z)X_\ell(s)-\lfloor Nt\rfloor C(z,s)\right)\,dz\,ds\,dx\end{aligned}$$

and

$$
\begin{aligned}
A^{(2)}_{\lfloor Nt\rfloor} &= \lfloor Nt\rfloor \hat{c}_{j,N}\int_0^1\int_0^1\int_0^1 v_j(z)\sum_{r=q+1}^{\infty}\sum_{n=1}^{\infty}\psi_{r,n}w_r(x)v_n(s)\left(\hat{w}_{i,N}(x)-\hat{d}_{i,N}w_i(x)\right)\\
&\qquad \times C(z,s)\,dz\,ds\,dx\\
&= \lfloor Nt\rfloor \hat{c}_{j,N}\int_0^1\int_0^1 \lambda_j v_j(s)\sum_{r=q+1}^{\infty}\sum_{n=1}^{\infty}\psi_{r,n}w_r(x)v_n(s)\left(\hat{w}_{i,N}(x)-\hat{d}_{i,N}w_i(x)\right)\,ds\,dx\\
&= \lfloor Nt\rfloor \hat{c}_{j,N}\int_0^1 \lambda_j \sum_{r=q+1}^{\infty}\psi_{r,j}w_r(x)\left(\hat{w}_{i,N}(x)-\hat{d}_{i,N}w_i(x)\right)\,dx\\
&= \lfloor Nt\rfloor R^{(1)}_N(i,j),
\end{aligned}
$$

where we used that the v_j's are orthonormal eigenfunctions of C. Applying again (5.20) and (5.25) we conclude

$$
\begin{aligned}
\sup_{t\in[0,1]}\left|A^{(1)}_{\lfloor Nt\rfloor}\right| &\le \left\|v_j(z)\sum_{r=q+1}^{\infty}\sum_{n=1}^{\infty}\psi_{r,n}w_r(x)v_n(s)\right\|\left\|\hat{w}_{i,N}(x)-\hat{d}_{i,N}w_i(x)\right\|\\
&\qquad \times \sup_{t\in[0,1]}\left\|\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(z)X_\ell(s)-\lfloor Nt\rfloor C(z,s)\right\|\\
&= O(1)O_P\left(N^{-1/2}\right)O_P\left(N^{1/2}\log N\right).
\end{aligned}
$$

Finally, using Theorems 5.5 and 5.6, we obtain that

$$
\begin{aligned}
&\sup_{t\in[0,1]}\left|\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,\hat{v}_{j,N}-\hat{c}_{j,N}v_j\rangle\langle\eta_{\ell,1},\hat{w}_{i,N}-\hat{d}_{i,N}w_i\rangle\right|\\
&= \sup_{t\in[0,1]}\left|\sum_{\ell=1}^{\lfloor Nt\rfloor}\int_0^1\int_0^1 X_\ell(z)\left(\hat{v}_{j,N}(z)-\hat{c}_{j,N}v_j(z)\right)\sum_{r=q+1}^{\infty}\sum_{n=1}^{\infty}\psi_{r,n}w_r(x)\right.\\
&\qquad \left.\times\int_0^1 v_n(s)X_\ell(s)\left(\hat{w}_{i,N}(x)-\hat{d}_{i,N}w_i(x)\right)\,dz\,ds\,dx\right|\\
&= \sup_{t\in[0,1]}\left|\int_0^1\int_0^1\int_0^1\left(\hat{v}_{j,N}(z)-\hat{c}_{j,N}v_j(z)\right)\sum_{r=q+1}^{\infty}\sum_{n=1}^{\infty}\psi_{r,n}w_r(x)v_n(s)\left(\hat{w}_{i,N}(x)-\hat{d}_{i,N}w_i(x)\right)\right.\\
&\qquad \left.\times\sum_{\ell}^{\lfloor Nt\rfloor}X_\ell(z)X_\ell(s)\,dz\,ds\,dx\right|\\
&\le ||\hat{v}_{j,N}(z)-\hat{c}_{j,N}v_j(z)||\left\|\sum_{r=q+1}^{\infty}\sum_{n=1}^{\infty}\psi_{r,n}w_r(x)v_n(s)\right\|\left\|\hat{w}_{i,N}(x)-\hat{d}_{i,N}w_i(x)\right\|\\
&\qquad \times\sup_{t\in[0,1]}\left\|\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(z)X_\ell(s)\right\|\\
&= O_P\left(N^{-1/2}\right)O(1)O_P\left(N^{-1/2}\right)O_P\left(N\right).
\end{aligned}
$$

□

Lemma 5.5. *If Assumptions 5.1–5.6 hold, then we have*

$$\sup_{t\in[0,1]}\left|\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,\hat{v}_{j,N}\rangle\langle\eta_{\ell,2},\hat{w}_{i,N}\rangle-\left(\hat{c}_{j,N}\hat{d}_{i,N}T^{(2)}_{\lfloor Nt\rfloor}(i,j)+\lfloor Nt\rfloor R^{(2)}_N(i,j)\right)\right|=O_P\left(\log N\right),$$

where

$$T^{(2)}_{\lfloor Nt\rfloor}(i,j)=\sum_{\ell=1}^{\lfloor Nt\rfloor}\int_0^1\int_0^1\left(X_\ell(s)X_\ell(z)-C(z,s)\right)\sum_{r=p+1}^{\infty}\psi_{ir}v_r(s)v_j(z)\,dz\,ds.$$

Proof. First we write

$$\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,\hat{v}_{j,N}\rangle\langle\eta_{\ell,2},\hat{w}_{i,N}\rangle=A^{(3)}_{\lfloor Nt\rfloor}+A^{(4)}_{\lfloor Nt\rfloor}+A^{(5)}_{\lfloor Nt\rfloor}+A^{(6)}_{\lfloor Nt\rfloor},$$

where

$$\begin{aligned}
A^{(3)}_{\lfloor Nt\rfloor}&=\hat{c}_{j,N}\hat{d}_{i,N}\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,v_j\rangle\langle\eta_{\ell,2},w_i\rangle,\\
A^{(4)}_{\lfloor Nt\rfloor}&=\hat{c}_{j,N}\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,v_j\rangle\langle\eta_{\ell,2},\hat{w}_{i,N}-\hat{d}_{i,N}w_i\rangle,\\
A^{(5)}_{\lfloor Nt\rfloor}&=\hat{d}_{i,N}\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,\hat{v}_{j,N}-\hat{c}_{j,N}v_j\rangle\langle\eta_{\ell,2},w_i\rangle,\\
A^{(6)}_{\lfloor Nt\rfloor}&=\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,\hat{v}_{j,N}-\hat{c}_{j,N}v_j\rangle\langle\eta_{\ell,2},\hat{w}_{i,N}-\hat{d}_{i,N}w_i\rangle.
\end{aligned}$$

The orthonormality of $\{w_i,\ 1\le i<\infty\}$ shows that for all $1\le i\le q$

$$\begin{aligned}
\langle\eta_{\ell,2},w_i\rangle&=\int_0^1\sum_{r=1}^{q}\sum_{n=p+1}^{\infty}\psi_{r,n}w_r(x)\left\{\int_0^1 v_n(s)X_\ell(s)\,ds\right\}w_i(x)\,dx\\
&=\sum_{n=p+1}^{\infty}\psi_{i,n}\int_0^1 v_n(s)X_\ell(s)\,ds.
\end{aligned}$$

Therefore, using again that the v_j's are orthonormal eigenfunctions of C we have

$$
\begin{aligned}
A_{\lfloor Nt \rfloor}^{(3)} &= \hat{c}_{j,N}\hat{d}_{i,N} \sum_{\ell=1}^{\lfloor Nt \rfloor} \int_0^1 X_\ell(z) v_j(z)\, dz \sum_{n=p+1}^{\infty} \psi_{i,n} \int_0^1 v_n(s) X_\ell(s)\, ds \\
&= \hat{c}_{j,N}\hat{d}_{i,N} \sum_{n=p+1}^{\infty} \psi_{i,n} \int_0^1 \int_0^1 v_j(z) v_n(s) \sum_{\ell=1}^{\lfloor Nt \rfloor} X_\ell(z) X_\ell(s)\, ds\, dz \\
&= \hat{c}_{j,N}\hat{d}_{i,N} \sum_{n=p+1}^{\infty} \psi_{i,n} \int_0^1 \int_0^1 v_j(z) v_n(s) \left(\sum_{\ell=1}^{\lfloor Nt \rfloor} X_\ell(z) X_\ell(s) - \lfloor Nt \rfloor C(s,z) \right) ds\, dz \\
&\quad + \hat{c}_{j,N}\hat{d}_{i,N} \lfloor Nt \rfloor \sum_{n=p+1}^{\infty} \psi_{i,n} \int_0^1 \int_0^1 v_i(z) v_n(s) C(s,z)\, ds\, dz \\
&= \hat{c}_{j,N}\hat{d}_{i,N} \sum_{n=p+1}^{\infty} \psi_{i,n} \int_0^1 \int_0^1 v_j(z) v_n(s) \left(\sum_{\ell=1}^{\lfloor Nt \rfloor} X_\ell(z) X_\ell(s) - \lfloor Nt \rfloor C(s,z) \right) ds\, dz \\
&= \hat{c}_{j,N}\hat{d}_{i,N} T_{\lfloor Nt \rfloor}^{(2)}(i,j).
\end{aligned}
$$

We decompose $A_{\lfloor Nt \rfloor}^{(4)}$ as

$$
\begin{aligned}
A_{\lfloor Nt \rfloor}^{(4)} &= \hat{c}_{j,N} \sum_{\ell=1}^{\lfloor Nt \rfloor} \int_0^1 \int_0^1 X_\ell(z) v_j(z) \left(\hat{w}_{i,N}(x) - \hat{d}_{i,N} w_i(x) \right) \sum_{r=1}^{q} \sum_{n=p+1}^{\infty} \psi_{r,n} w_r(x) \\
&\qquad \times \int_0^1 v_n(s) X_\ell(s)\, ds\, dz\, dx \\
&= \hat{c}_{j,N} \int_0^1 \int_0^1 \int_0^1 v_j(z) \left(\hat{w}_{i,N}(x) - \hat{d}_{i,N} w_i(x) \right) \sum_{r=1}^{q} \sum_{n=p+1}^{\infty} \psi_{r,n} w_r(x) v_n(s) \\
&\qquad \times \sum_{\ell=1}^{\lfloor Nt \rfloor} X_\ell(s) X_\ell(z)\, ds\, dz\, dx \\
&= A_{\lfloor Nt \rfloor,1}^{(4)} + A_{\lfloor Nt \rfloor,2}^{(4)},
\end{aligned}
$$

where

$$
\begin{aligned}
A_{\lfloor Nt \rfloor,1}^{(4)} &= \hat{c}_{j,N} \int_0^1 \int_0^1 \int_0^1 v_j(z) \left(\hat{w}_{i,N}(x) - \hat{d}_{i,N} w_i(x) \right) \sum_{r=1}^{q} \sum_{n=p+1}^{\infty} \psi_{r,n} w_r(x) v_n(s) \\
&\qquad \times \left(\sum_{\ell=1}^{\lfloor Nt \rfloor} X_\ell(s) X_\ell(z) - \lfloor Nt \rfloor C(s,z) \right) ds\, dz\, dx
\end{aligned}
$$

and

$$
\begin{aligned}
A^{(4)}_{\lfloor Nt \rfloor,2} &= \hat{c}_{j,N} \lfloor Nt \rfloor \int_0^1 \int_0^1 \int_0^1 v_j(z) \left(\hat{w}_{i,N}(x) - \hat{d}_{i,N} w_i(x) \right) \\
&\quad \times \sum_{r=1}^{q} \sum_{n=p+1}^{\infty} \psi_{r,n} w_r(x) v_n(s) C(s,z) \, ds \, dz \, dx \\
&= \hat{c}_{j,N} \lfloor Nt \rfloor \int_0^1 \int_0^1 \lambda_j v_j(s) \left(\hat{w}_{i,N}(x) - \hat{d}_{i,N} w_i(x) \right) \sum_{r=1}^{q} \sum_{n=p+1}^{\infty} \psi_{r,n} w_r(x) v_n(s) \, ds \, dx \\
&= 0,
\end{aligned}
$$

using again that the v_j's are eigenfunctions of C. Therefore we obtain

$$
\begin{aligned}
\sup_{t \in [0,1]} \left| A^{(4)}_{\lfloor Nt \rfloor} \right| &= \sup_{t \in [0,1]} \left| A^{(4)}_{\lfloor Nt \rfloor,1} \right| \\
&\leq \left\| \hat{w}_{i,N}(x) - \hat{d}_{i,N} w_i(x) \right\| \left\| v_j(z) \sum_{r=1}^{q} \sum_{n=p+1}^{\infty} \psi_{r,n} w_r(x) v_n(s) \right\| \\
&\quad \times \sup_{t \in [0,1]} \left\| \sum_{\ell=1}^{\lfloor Nt \rfloor} X_\ell(s) X_\ell(z) - \lfloor Nt \rfloor C(s,z) \right\| \\
&= O_P\left(N^{-1/2}\right) O_P(1) \, O_P\left(N^{1/2} \log N\right).
\end{aligned}
$$

Similar arguments give

$$
\begin{aligned}
A^{(5)}_{\lfloor Nt \rfloor} &= \hat{d}_{i,N} \sum_{\ell=1}^{\lfloor Nt \rfloor} \int_0^1 \int_0^1 X_\ell(z) \left(\hat{v}_{j,N}(z) - \hat{c}_{j,N} v_j(z) \right) w_i(x) \sum_{r=1}^{q} \sum_{n=p+1}^{\infty} \psi_{r,n} w_r(x) \\
&\quad \times \int_0^1 v_n(s) X_\ell(s) \, ds \, dz \, dx \\
&= \hat{d}_{i,N} \int_0^1 \int_0^1 \int_0^1 \left(\hat{v}_{j,N}(z) - \hat{c}_{j,N} v_j(z) \right) w_i(x) \sum_{r=1}^{q} \sum_{n=p+1}^{\infty} \psi_{r,n} w_r(x) v_n(s) \\
&\quad \times \sum_{\ell=1}^{\lfloor Nt \rfloor} X_\ell(s) X_\ell(z) \, ds \, dz \, dx \\
&= A^{(5)}_{\lfloor Nt \rfloor,1} + A^{(5)}_{\lfloor Nt \rfloor,2},
\end{aligned}
$$

where

$$
\begin{aligned}
A^{(5)}_{\lfloor Nt\rfloor,1} &= \hat{d}_{i,N}\int_0^1\int_0^1\int_0^1 \left(\hat{v}_{j,N}(z)-\hat{c}_{j,N}v_j(z)\right) w_i(x)\sum_{r=1}^{q}\sum_{n=p+1}^{\infty}\psi_{r,n}w_r(x)v_n(s)\\
&\qquad\times\left(\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(s)X_\ell(z)-\lfloor Nt\rfloor C(s,z)\right)ds\,dz\,dx\\
&= \hat{d}_{i,N}\int_0^1\int_0^1 \left(\hat{v}_{j,N}(z)-\hat{c}_{j,N}v_j(z)\right)\sum_{n=p+1}^{\infty}\psi_{i,n}v_n(s)\\
&\qquad\times\left(\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(s)X_\ell(z)-\lfloor Nt\rfloor C(s,z)\right)ds\,dz
\end{aligned}
$$

and

$$
\begin{aligned}
A^{(5)}_{\lfloor Nt\rfloor,2} &= \hat{d}_{i,N}\lfloor Nt\rfloor\int_0^1\int_0^1\int_0^1 \left(\hat{v}_{j,N}(z)-\hat{c}_{j,N}v_j(z)\right) w_i(x)\sum_{r=1}^{q}\sum_{n=p+1}^{\infty}\psi_{r,n}w_r(x)v_n(s)\\
&\qquad\times C(s,z)\,ds\,dz\,dx\\
&= \hat{d}_{i,N}\lfloor Nt\rfloor\int_0^1\int_0^1 \left(\hat{v}_{j,N}(z)-\hat{c}_{j,N}v_j(z)\right) w_i(x)\sum_{r=1}^{q}\sum_{n=p+1}^{\infty}\psi_{r,n}w_r(x)\lambda_n v_n(z)\,dz\,dx\\
&= \hat{d}_{i,N}\lfloor Nt\rfloor\int_0^1 \left(\hat{v}_{j,N}(z)-\hat{c}_{j,N}v_j(z)\right)\sum_{n=p+1}^{\infty}\psi_{i,n}\lambda_n v_n(z)\,dz\\
&= \lfloor Nt\rfloor R^{(2)}_N(i,j).
\end{aligned}
$$

Repeating our previous arguments we get that

$$
\begin{aligned}
\sup_{t\in[0,1]}\left|A^{(5)}_{\lfloor Nt\rfloor,1}\right| &\le \left\|\hat{v}_{j,N}(z)-\hat{c}_{j,N}v_j(z)\right\|\left\|\sum_{n=p+1}^{\infty}\psi_{i,n}v_n(s)\right\|\sup_{t\in[0,1]}\left\|\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(s)X_\ell(z)-\lfloor Nt\rfloor C(s,z)\right\|\\
&= O_P\left(N^{-1/2}\right)O(1)O_P\left(N^{1/2}\log N\right).
\end{aligned}
$$

Similarly, using the Cauchy-Schwarz inequality with (5.17) and Theorem 5.5, we conclude that

$$
\begin{aligned}
&\sup_{t\in[0,1]} |A^{(6)}_{\lfloor Nt\rfloor}| \\
&\quad = \sup_{t\in[0,1]} \Bigg| \sum_{\ell=1}^{\lfloor Nt\rfloor} X_\ell(z) \int_0^1 \int_0^1 \left(\hat{v}_{j,N}(z) - \hat{c}_{j,N} v_j(z)\right) \left(\hat{w}_{i,N}(x) - \hat{d}_{i,N} w_i(x)\right) \\
&\qquad \times \sum_{r=1}^{q} \sum_{n=p+1}^{\infty} \psi_{r,n} w_r(x) \int_0^1 v_n(s) X_\ell(s)\, ds\, dx\, dz \Bigg| \\
&\quad = \sup_{t\in[0,1]} \Bigg| \int_0^1 \int_0^1 \int_0^1 \left(\hat{v}_{j,N}(z) - \hat{c}_{j,N} v_j(z)\right) \left(\hat{w}_{i,N}(x) - \hat{d}_{i,N} w_i(x)\right) \\
&\qquad \times \sum_{r=1}^{q} \sum_{n=p+1}^{\infty} \psi_{r,n} w_r(x) v_n(s) \sum_{\ell=1}^{\lfloor Nt\rfloor} X_\ell(z) X_\ell(s)\, ds\, dx\, dz \Bigg| \\
&\quad \le \|\hat{w}_{i,N}(x) - \hat{d}_{i,N} w_i(x)\| \sup_{t\in[0,1]} \left\| \sum_{\ell=1}^{\lfloor Nt\rfloor} X_\ell(z) X_\ell(s) \right\| \\
&\qquad \times \|\hat{v}_{j,N}(z) - \hat{c}_{j,N} v_j(z)\| \left\| \sum_{r=1}^{q} \sum_{n=p+1}^{\infty} \psi_{r,n} w_r(x) v_n(s) \right\| \\
&\quad = O_P\left(N^{-1/2}\right) O_P\left(N\right) O_P\left(N^{-1/2}\right) O(1),
\end{aligned}
$$

completing the proof of the lemma. □

Lemma 5.6. *If Assumptions 5.1–5.6 hold, then we have*

$$
\sup_{t\in[0,1]} \left| \sum_{\ell=1}^{\lfloor Nt\rfloor} \langle X_\ell, \hat{v}_{j,N}\rangle \langle \eta_{\ell,3}, \hat{w}_{i,N}\rangle - \lfloor Nt\rfloor R_N^{(3)}(i,j) \right| = O_P\left(\log N\right).
$$

Proof. We write

$$
\sum_{\ell=1}^{\lfloor Nt\rfloor} \langle X_\ell, \hat{v}_{j,N}\rangle \langle \eta_{\ell,3},\ \hat{w}_{i,N}\rangle = A^{(7)}_{\lfloor Nt\rfloor} + A^{(8)}_{\lfloor Nt\rfloor} + A^{(9)}_{\lfloor Nt\rfloor},
$$

where

$$
\begin{aligned}
A^{(7)}_{\lfloor Nt\rfloor} &= \sum_{\ell=1}^{\lfloor Nt\rfloor} \langle X_\ell, \hat{v}_{j,N} - \hat{c}_{j,N} v_j\rangle \langle \eta_{\ell,3}, \hat{w}_{i,N}\rangle, \\
A^{(8)}_{\lfloor Nt\rfloor} &= \hat{c}_{j,N} \sum_{\ell=1}^{\lfloor Nt\rfloor} \langle X_\ell, v_j\rangle \langle \eta_{\ell,3}, \hat{w}_{i,N} - \hat{d}_{i,N} w_i\rangle, \\
A^{(9)}_{\lfloor Nt\rfloor} &= \hat{c}_{j,N} \hat{d}_{i,N} \sum_{\ell=1}^{\lfloor Nt\rfloor} \langle X_\ell,\ v_j\rangle \langle \eta_{\ell,3}, w_i\rangle.
\end{aligned}
$$

Theorems 5.4 and 5.5 imply that

$$
\begin{aligned}
\sup_{t\in[0,1]}\left|A^{(7)}_{\lfloor Nt\rfloor}\right| &= \sup_{t\in[0,1]}\left|\int_0^1\int_0^1\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(z)\left(\hat{v}_{j,N}(z)-\hat{c}_{j,N}v_j(z)\right)\hat{w}_{i,N}(x)\sum_{r=1}^{q}\sum_{n=1}^{p}\psi_{r,n}\hat{c}_{n,N}w_r(x)\right.\\
&\qquad\left.\times\int_0^1\left(\hat{c}_{n,N}v_n(s)-\hat{v}_{n,N}(s)\right)X_\ell(s)\,ds\,dz\,dx\right|\\
&\le\sum_{r=1}^{q}\sum_{n=1}^{p}\sup_{t\in[0,1]}\left\|\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(s)X_\ell(z)\right\|\left\|\hat{v}_{j,N}(z)-\hat{c}_{j,N}v_j(z)\right\|\\
&\qquad\times\left\|\hat{w}_{i,N}(x)\psi_{r,n}\hat{c}_{n,N}w_r(x)\right\|\left\|\hat{c}_{n,N}v_n(s)-\hat{v}_{n,N}(s)\right\|\\
&=O_P\left(N\right)O_P\left(N^{-1/2}\right)O(1)O_P\left(N^{-1/2}\right),
\end{aligned}
$$

and similarly

$$
\sup_{t\in[0,1]}\left|A^{(8)}_{\lfloor Nt\rfloor}\right| = O_P\left(1\right).
$$

Next we observe that

$$
\begin{aligned}
\sup_{t\in[0,1]}\left|A^{(9)}_{\lfloor Nt\rfloor}\right| &= \hat{c}_{j,N}\hat{d}_{i,N}\sum_{\ell=1}^{\lfloor Nt\rfloor}\int_0^1\int_0^1 X_\ell(z)v_j(z)w_i(x)\sum_{r=1}^{q}\sum_{n=1}^{p}\psi_{r,n}\hat{c}_{n,N}w_r(x)\\
&\qquad\times\int_0^1\left(\hat{c}_{n,N}v_n(s)-\hat{v}_{n,N}(s)\right)X_\ell(s)\,ds\,dz\,dx\\
&=\hat{c}_{j,N}\hat{d}_{i,N}\int_0^1\int_0^1\int_0^1\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(z)X_\ell(s)v_j(z)w_i(x)\sum_{r=1}^{q}\sum_{n=1}^{p}\psi_{r,n}\hat{c}_{n,N}w_r(x)\\
&\qquad\times\left(\hat{c}_{n,N}v_n(s)-\hat{v}_{n,N}(s)\right)\,ds\,dz\,dx\\
&=A^{(9)}_{\lfloor Nt\rfloor,1}+A^{(9)}_{\lfloor Nt\rfloor,2},
\end{aligned}
$$

where

$$
\begin{aligned}
A^{(9)}_{\lfloor Nt\rfloor,1} &= \hat{c}_{j,N}\hat{d}_{i,N}\lfloor Nt\rfloor\int_0^1\int_0^1\int_0^1 C(z,s)v_j(z)w_i(x)\sum_{r=1}^{q}\sum_{n=1}^{p}\psi_{r,n}\hat{c}_{n,N}w_r(x)\\
&\qquad\times\left(\hat{c}_{n,N}v_n(s)-\hat{v}_{n,N}(s)\right)\,ds\,dz\,dx\\
&=\hat{c}_{j,N}\hat{d}_{i,N}\lfloor Nt\rfloor\int_0^1\int_0^1\lambda_j v_j(s)w_i(x)\sum_{r=1}^{q}\sum_{n=1}^{p}\psi_{r,n}\hat{c}_{n,N}w_r(x)\left(\hat{c}_{n,N}v_n(s)-\hat{v}_{n,N}(s)\right)\,ds\,dx\\
&=\hat{c}_{j,N}\hat{d}_{i,N}\lfloor Nt\rfloor\int_0^1\lambda_j v_j(s)\sum_{n=1}^{p}\psi_{i,n}\hat{c}_{n,N}\left(\hat{c}_{n,N}v_n(s)-\hat{v}_{n,N}(s)\right)\,ds\\
&=\lfloor Nt\rfloor R^{(3)}_N(i,j)
\end{aligned}
$$

and

$$
\begin{aligned}
A^{(9)}_{\lfloor Nt\rfloor,2} &= \hat{c}_{j,N}\hat{d}_{i,N}\int_0^1\int_0^1\int_0^1\left(\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(z)X_\ell(s)-\lfloor Nt\rfloor C(z,s)\right)v_j(z)w_i(x)\\
&\qquad\times\sum_{r=1}^{q}\sum_{n=1}^{p}\psi_{r,n}\hat{c}_{n,N}w_r(x)\left(\hat{c}_{n,N}v_n(s)-\hat{v}_{n,N}(s)\right)\,ds\,dz\,dx.
\end{aligned}
$$

Using Theorems 5.5 and 5.6 again, we obtain that

$$\sup_{t\in[0,1]} |A^{(9)}_{\lfloor Nt\rfloor,2}| = O_P(\log N).$$

This completes the proof. □

Lemma 5.7. *If Assumptions 5.1–5.6 hold, then we have*

$$\sup_{t\in[0,1]} \left| \sum_{\ell=1}^{\lfloor Nt\rfloor} \langle X_\ell, \hat{v}_{j,N}\rangle\langle \eta_{\ell,4}, \hat{w}_{i,N}\rangle - \lfloor Nt\rfloor R_N^{(4)}(i,j) \right| = O_P(\log N).$$

Proof. Following the proofs of the previous lemmas we write

$$\sum_{\ell=1}^{\lfloor Nt\rfloor} \langle X_\ell, \hat{v}_{j,N}\rangle\langle \eta_{\ell,4}, \hat{w}_{i,N}\rangle = A^{(10)}_{\lfloor Nt\rfloor} + A^{(11)}_{\lfloor Nt\rfloor} + A^{(12)}_{\lfloor Nt\rfloor} + A^{(13)}_{\lfloor Nt\rfloor},$$

where

$$\begin{aligned}
A^{(10)}_{\lfloor Nt\rfloor} &= \sum_{\ell=1}^{\lfloor Nt\rfloor} \langle X_\ell, \hat{v}_{j,N} - \hat{c}_{j,N} v_j\rangle\langle \eta_{\ell,4}, \hat{w}_{i,N} - \hat{d}_{i,N} w_i\rangle,\\
A^{(11)}_{\lfloor Nt\rfloor} &= \hat{c}_{j,N}\hat{d}_{i,N} \sum_{\ell=1}^{\lfloor Nt\rfloor} \langle X_\ell,\ v_j\rangle\langle \eta_{\ell,4}, w_i\rangle,\\
A^{(12)}_{\lfloor Nt\rfloor} &= \hat{d}_{i,N} \sum_{\ell=1}^{\lfloor Nt\rfloor} \langle X_\ell, \hat{v}_{j,N} - \hat{c}_{j,N} v_j\rangle\langle \eta_{\ell,4}, w_i\rangle,\\
A^{(13)}_{\lfloor Nt\rfloor} &= \hat{c}_{j,N} \sum_{\ell=1}^{\lfloor Nt\rfloor} \langle X_\ell, v_j\rangle\langle \eta_{\ell,4}, \hat{w}_{i,N} - \hat{d}_{i,N} w_i\rangle.
\end{aligned}$$

Repeating the arguments used in the proofs of Lemmas 5.4 and 5.5, one can show that

$$\begin{aligned}
\sup_{t\in[0,1]} \left|A^{(10)}_{\lfloor Nt\rfloor}\right| &= O_P(1),\\
\sup_{t\in[0,1]} \left|A^{(12)}_{\lfloor Nt\rfloor}\right| &= O_P(1),\\
\sup_{t\in[0,1]} \left|A^{(13)}_{\lfloor Nt\rfloor}\right| &= O_P(1).
\end{aligned}$$

Elementary arguments give

$$
\begin{aligned}
A^{(11)}_{\lfloor Nt\rfloor} &= \hat{c}_{j,N}\hat{d}_{i,N}\sum_{\ell=1}^{\lfloor Nt\rfloor}\int_0^1\int_0^1 X_\ell(z)v_j(z)w_i(x)\sum_{r=1}^{q}\sum_{n=1}^{p}\hat{d}_{r,N}\psi_{r,n}\hat{c}_{n,N}\left(\hat{d}_{r,N}w_r(x)-\hat{w}_{r,N}(x)\right)\\
&\qquad\times\int_0^1 \hat{c}_{n,N}v_n(s)X_\ell(s)\,ds\,dz\,dx\\
&= \hat{c}_{j,N}\hat{d}_{i,N}\int_0^1\int_0^1\int_0^1\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(z)X_\ell(s)v_j(z)w_i(x)\sum_{r=1}^{q}\sum_{n=1}^{p}\hat{d}_{r,N}\psi_{r,n}\hat{c}_{n,N}\\
&\qquad\times\left(\hat{d}_{r,N}w_r(x)-\hat{w}_{r,N}(x)\right)\hat{c}_{n,N}v_n(s)\,ds\,dz\,dx\\
&= A^{(11)}_{\lfloor Nt\rfloor,1}+A^{(11)}_{\lfloor Nt\rfloor,2},
\end{aligned}
$$

where

$$
\begin{aligned}
A^{(11)}_{\lfloor Nt\rfloor,2} = \hat{c}_{j,N}\hat{d}_{i,N}\int_0^1\int_0^1\int_0^1&\left(\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(z)X_\ell(s)-\lfloor Nt\rfloor C(z,s)\right)v_j(z)w_i(x)\\
&\times\sum_{r=1}^{q}\sum_{n=1}^{p}\hat{d}_{r,N}\psi_{r,n}\hat{c}_{n,N}\left(\hat{d}_{r,N}w_r(x)-\hat{w}_{r,N}(x)\right)\hat{c}_{n,N}v_n(s)\,ds\,dz\,dx,
\end{aligned}
$$

and

$$
\begin{aligned}
A^{(11)}_{\lfloor Nt\rfloor,1} &= \lfloor Nt\rfloor\hat{c}_{j,N}\hat{d}_{i,N}\int_0^1\int_0^1\int_0^1 C(z,s)v_j(z)w_i(x)\sum_{r=1}^{q}\sum_{n=1}^{p}\hat{d}_{r,N}\psi_{r,n}\hat{c}_{n,N}\\
&\qquad\times\left(\hat{d}_{r,N}w_r(x)-\hat{w}_{r,N}(x)\right)\hat{c}_{n,N}v_n(s)\,ds\,dz\,dx\\
&= \lfloor Nt\rfloor\hat{c}_{j,N}\hat{d}_{i,N}\int_0^1\int_0^1\lambda_j v_j(s)w_i(x)\sum_{r=1}^{q}\sum_{n=1}^{p}\hat{d}_{r,N}\psi_{r,n}\hat{c}_{n,N}\\
&\qquad\times\left(\hat{d}_{r,N}w_r(x)-\hat{w}_{r,N}(x)\right)\hat{c}_{n,N}v_n(s)\,ds\,dx\\
&= \lfloor Nt\rfloor\hat{c}_{j,N}\hat{d}_{i,N}\lambda_j\int_0^1 w_i(x)\sum_{r=1}^{q}\hat{d}_{r,N}\psi_{r,j}\hat{c}_{j,N}\left(\hat{d}_{r,N}w_r(x)-\hat{w}_{r,N}(x)\right)\hat{c}_{j,N}\,dx\\
&= \lfloor Nt\rfloor\hat{c}_{j,N}\hat{d}_{i,N}\lambda_j\sum_{r=1}^{q}\hat{d}_{r,N}\psi_{r,j}\int_0^1 w_i(x)\left(\hat{d}_{r,N}w_r(x)-\hat{w}_{r,N}(x)\right)\,dx\\
&= \lfloor Nt\rfloor R^{(4)}_N(i,j).
\end{aligned}
$$

Using Theorems 5.5 and 5.6 again, we conclude that

$$
\sup_{t\in[0,1]}\left|A^{(11)}_{\lfloor Nt\rfloor,2}\right| = O_P\left(\log N\right),
$$

completing the proof. □

Lemma 5.8. *If Assumptions 5.1–5.5 hold, then we have*

$$
\sup_{t\in[0,1]}\left|\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,\hat{v}_{j,N}\rangle\langle\eta_{\ell,5},\hat{w}_{i,N}\rangle\right| = O_P\left(1\right).
$$

Proof. It follows from Theorems 5.4 and 5.5 that

$$\begin{aligned}
&\sup_{t\in[0,1]}\left|\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,\hat v_{j,N}\rangle\langle \eta_{\ell,5},\hat w_{i,N}\rangle\right| \\
&=\sup_{t\in[0,1]}\left|\sum_{\ell=1}^{\lfloor Nt\rfloor}\int_0^1\int_0^1 X_\ell(z)\hat v_{j,N}(z)\hat w_{i,N}(x)\sum_{r=1}^{q}\sum_{n=1}^{p}\hat d_{r,N}\psi_{r,n}\hat c_{n,N}\left(\hat w_{r,N}(x)-\hat d_{r,N}w_r(x)\right)\right.\\
&\qquad\left.\times\int_0^1\left(\hat c_{n,N}v_n(s)-\hat v_{n,N}(s)\right)X_\ell(s)\,ds\,dz\,dx\right|\\
&\le\sum_{r=1}^{q}\sum_{n=1}^{p}|\psi_{r,n}|\sup_{t\in[0,1]}\left\|\sum_{\ell=1}^{\lfloor Nt\rfloor}X_\ell(z)X_\ell(s)\right\|\|\hat w_{i,N}(x)\|\,\|\hat v_{j,N}(z)\|\left\|\hat w_{r,N}(x)-\hat d_{r,N}w_r(x)\right\|\\
&\qquad\times\|\hat c_{n,N}v_n(s)-\hat v_{n,N}(s)\|\\
&=O_P\left(N\right)O_P\left(N^{-1/2}\right)O_P\left(N^{-1/2}\right).
\end{aligned}$$

□

Lemma 5.9. *If Assumptions 5.1–5.6 hold, then we have*

$$\sup_{t\in[0,1]}\left|\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,\hat v_{j,N}\rangle\langle\epsilon_\ell^{**},\hat w_{i,N}\rangle-\left(\lfloor Nt\rfloor R_N(i,j)+\hat c_{j,N}\hat d_{i,N}\sum_{\ell=1}^{\lfloor Nt\rfloor}\gamma_\ell(i,j)\right)\right|=O_P\left(\log N\right).$$

Proof. Combining Lemmas 5.3–5.8, we immediately see that

$$\sup_{t\in[0,1]}\left|\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,\hat v_{j,N}\rangle\langle\epsilon_\ell^{**},\hat w_{i,N}\rangle-\left(\lfloor Nt\rfloor R_N(i,j)+\hat c_{j,N}\hat d_{i,N}T_{\lfloor Nt\rfloor}(i,j)\right)\right|=O_P\left(\log N\right),$$

where

$$T_{\lfloor Nt\rfloor}(i,j)=T^{(1)}_{\lfloor Nt\rfloor}(i,j)+T^{(2)}_{\lfloor Nt\rfloor}(i,j).$$

Thus we need only to show that

$$T^{(2)}_{\lfloor Nt\rfloor}(i,j)=\sum_{\ell=1}^{\lfloor Nt\rfloor}\langle X_\ell,v_j\rangle\langle X_\ell,u_i\rangle.$$

However, using that the v_j's are orthogonal eigenfunctions of C, we get that

$$\int_0^1\int_0^1 C(z,s)v_j(z)\sum_{r=p+1}^{\infty}\psi_{i,r}v_r(s)\,ds\,dz=\lambda_j\int_0^1 v_j(s)\sum_{r=p+1}^{\infty}\psi_{i,r}v_r(s)\,ds\,dz=0,$$

completing the proof. □

Lemma 5.10. *If Assumptions 5.1–5.6 hold, then we have*

$$\left|\sqrt{N}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_N)\right|=O_P\left(1\right).$$

Proof. It is easy to see that

$$\sqrt{N}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_N) = -N^{-1/2}\left(\frac{\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N}{N}\right)^{-1}\hat{\mathbf{Z}}_N^T\hat{\boldsymbol{\Delta}}_N.$$

It follows from (5.33) that

$$\left|\left(\frac{\hat{\mathbf{Z}}_N^T\hat{\mathbf{Z}}_N}{N}\right)^{-1}\right| = O_P(1).$$

Lemma 5.9 and (5.35) yield that

$$\left|\hat{\mathbf{Z}}_N^T\hat{\boldsymbol{\Delta}}_N\right| \le \max_{1\le i\le q, 1\le j\le p}\left\{N|R_N(i,j)| + \left|\sum_{\ell=1}^N \gamma_\ell(i,j)\right|\right\} + O_P(\log N).$$

It follows from Theorem 5.5 that for all $1 \le i \le q, 1 \le j \le p$

$$N|R_N(i,j)| = O_P\left(N^{1/2}\right)$$

while Theorem 5.7 implies that

$$\left|\sum_{\ell=1}^N \gamma_\ell(i,j)\right| = O_P\left(N^{1/2}\right).$$

□

Lemma 5.11. *If Assumptions 5.1–5.6 hold, then we have*

$$\sup_{t\in[0,1]}\left|\tilde{\mathbf{V}}_N(t) - \boldsymbol{\zeta}_N\frac{1}{N^{1/2}}\left(\sum_{\ell=1}^{\lfloor Nt\rfloor}\boldsymbol{\gamma}_\ell - t\sum_{\ell=1}^N\boldsymbol{\gamma}_\ell\right)\right| = o_P(1).$$

Proof. Lemmas 5.2 and 5.10 and (5.29) imply that

$$\sup_{t\in[0,1]}\left|\tilde{\mathbf{V}}_N(t) - \frac{1}{N^{1/2}}\left(\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\boldsymbol{\Delta}}_{\lfloor Nt\rfloor} - t\hat{\mathbf{Z}}_N^T\hat{\boldsymbol{\Delta}}_N\right)\right| = o_P(1).$$

It also follows from Lemma 5.9 and (5.35)

$$\sup_{t\in[0,1]}\left|\hat{\mathbf{Z}}_{\lfloor Nt\rfloor}^T\hat{\boldsymbol{\Delta}}_{\lfloor Nt\rfloor} - \mathrm{vec}\left(\left\{\lfloor Nt\rfloor R_N(i,j) + \hat{c}_{j,N}\hat{d}_{i,N}\sum_{\ell=1}^{\lfloor Nt\rfloor}\gamma_\ell(i,j)\right\}^T\right)\right| = O_P(\log N),$$

and therefore the proof is complete. □

Now we have all the necessary tools to prove the main result.

Proof of Theorem 5.1. It follows from Lemma 5.11 and Theorem 5.7 that

$$\boldsymbol{\zeta}_N \tilde{\mathbf{V}}_N(t) \quad \stackrel{\mathcal{D}^{pq}[0,1]}{\longrightarrow} \quad \mathbf{W}_{\boldsymbol{\Sigma}}(t) - t\mathbf{W}_{\boldsymbol{\Sigma}}(1).$$

Next we observe that

$$\left\{\boldsymbol{\Sigma}^{-1/2}(\mathbf{W}_{\boldsymbol{\Sigma}}(t) - t\mathbf{W}_{\boldsymbol{\Sigma}}(1)), 0 \leq t \leq 1\right\} \stackrel{\mathcal{D}}{=} \{\boldsymbol{B}(t), 0 \leq t \leq 1\},$$

where $\boldsymbol{B}(t) = (B_1(t), \dots B_{pq}(t))^T$ and $B_1, \dots B_{pq}$ are independent, identically distributed Brownian bridges. Hence

$$(\boldsymbol{\zeta}_N \tilde{\mathbf{V}}_N(t))^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\zeta}_N \tilde{\mathbf{V}}_N(t)) \quad \stackrel{\mathcal{D}[0,1]}{\longrightarrow} \quad \sum_{\ell=1}^{pq} B_\ell^2(t).$$

Now, using Assumption 5.7 with Slutsky's lemma, the proof is complete. $\square$

5.6 Proof of Theorems 5.2 and 5.3

We can assume without loss of generality that $K(u) = 0$ if $|u| > 1$. Let m be a positive integer and define

$$\boldsymbol{\gamma}_\ell^{(m)} = \mathrm{vec}(\{\gamma_\ell^{(m)}(i,j), 1 \leq i \leq q, 1 \leq j \leq p\}^T),$$

where

$$\gamma_\ell^{(m)}(i,j) = \langle X_\ell^{(m)}, v_j \rangle \langle \epsilon_\ell^{(m)}, w_i \rangle + \langle X_\ell^{(m)}, v_j \rangle \langle X_\ell^{(m)}, u_i \rangle.$$

The long term covariance matrix associated with the stationary sequence $\{\boldsymbol{\gamma}_\ell^{(m)}, 1 \leq \ell < \infty\}$ is given by

$$\boldsymbol{\Sigma}^{(m)} = E\boldsymbol{\gamma}_1^{(m)}(\boldsymbol{\gamma}_1^{(m)})^T + \sum_{\ell=1}^{\infty} E\boldsymbol{\gamma}_1^{(m)}(\boldsymbol{\gamma}_{\ell+1}^{(m)})^T + \sum_{\ell=1}^{\infty} E\boldsymbol{\gamma}_{\ell+1}^{(m)}(\boldsymbol{\gamma}_1^{(m)})^T.$$

The corresponding Bartlett estimator is defined as

$$\tilde{\boldsymbol{\Sigma}}_N^{(m)} = \sum_{k=-(N-1)}^{N-1} K(k/B_N)\boldsymbol{\phi}_{k,N}^{(m)},$$

where

$$\boldsymbol{\phi}_{k,N}^{(m)} = \frac{1}{N} \sum_{\ell=\max(1,1-k)}^{\min(N,N-k)} \boldsymbol{\gamma}_\ell^{(m)}(\boldsymbol{\gamma}_{\ell+k}^{(m)})^T$$

are the sample covariances of lag k. Since K is symmetric, $K(0) = 1$ and $K(u) = 0$ outside $[-1, 1]$ we have that

$$\tilde{\boldsymbol{\Sigma}}_N^{(m)} = \boldsymbol{\phi}_{0,N}^{(m)} + \sum_{k=1}^{B_N} K(k/B_N)\boldsymbol{\phi}_{k,N}^{(m)} + \sum_{k=1}^{B_N} K(k/B_N)(\boldsymbol{\phi}_{k,N}^{(m)})^T$$

for all sufficiently large N.

We start with the consistency of $\tilde{\boldsymbol{\Sigma}}_N^{(m)}$.

Lemma 5.12. *If Assumptions 5.8 and 5.9 are satisfied, then we have for every m*

$$\tilde{\boldsymbol{\Sigma}}_N^{(m)} \xrightarrow{P} \boldsymbol{\Sigma}^{(m)},$$

as $N \to \infty$.

Proof. Since the sequence $\boldsymbol{\gamma}_\ell^{(m)}$ is m-dependent we have that

$$\boldsymbol{\Sigma}^{(m)} = E\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^T + \sum_{\ell=1}^{m} E\boldsymbol{\gamma}_1\boldsymbol{\gamma}_{\ell+1}^T + \sum_{\ell=1}^{m} E\boldsymbol{\gamma}_{\ell+1}\boldsymbol{\gamma}_1^T.$$

It follows from the ergodic theorem that for any fixed k and m

$$\boldsymbol{\phi}_{k,N}^{(m)} \xrightarrow{P} E\boldsymbol{\gamma}_1^{(m)}(\boldsymbol{\gamma}_{1+k}^{(m)})^T.$$

So using Assumptions 5.8(i), 5.8(ii) and 5.9 we get that

$$\boldsymbol{\phi}_{0,N}^{(m)} + \sum_{k=1}^{m} K(k/B_N)\boldsymbol{\phi}_{k,N}^{(m)} + \sum_{k=1}^{m} K(k/B_N)(\boldsymbol{\phi}_{k,N}^{(m)})^T \xrightarrow{P} E\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^T + \sum_{\ell=1}^{m} E\boldsymbol{\gamma}_1\boldsymbol{\gamma}_{\ell+1}^T + \sum_{\ell=1}^{m} E\boldsymbol{\gamma}_{\ell+1}\boldsymbol{\gamma}_1^T.$$

Lemma 5.12 is proven if we show that

$$\sum_{k=m+1}^{B_N} K(k/B_N)\boldsymbol{\phi}_{k,N}^{(m)} \xrightarrow{P} 0 \tag{5.37}$$

and

$$\sum_{k=m+1}^{B_N} K(k/B_N)(\boldsymbol{\phi}_{k,N}^{(m)})^T \xrightarrow{P} 0. \tag{5.38}$$

Clearly, it is enough to prove (5.37).

Let

$$\mathbf{G}_N^{(m)} = \sum_{k=m+1}^{B_N} K(k/B_N)\boldsymbol{\phi}_{k,N}^{(m)}.$$

Elementary arguments show that

$$\begin{aligned}
\mathbf{G}_N^{(m)} &= \sum_{k=m+1}^{B_N} K(k/B_N)\boldsymbol{\phi}_{k,N}^{(m)} \\
&= \sum_{k=m+1}^{B_N} K(k/B_N)\frac{1}{N}\sum_{\ell=1}^{N-k} \boldsymbol{\gamma}_\ell^{(m)}\left(\boldsymbol{\gamma}_{\ell+k}^{(m)}\right)^T \\
&= \sum_{\ell=1}^{N-(m+1)} \boldsymbol{\gamma}_\ell^{(m)}\mathbf{H}_{\ell,N}^{(m)},
\end{aligned}$$

where

$$\mathbf{H}_{\ell,N}^{(m)} = \sum_{k=m+1}^{\min(N-\ell,B_N)} \frac{K(k/B_N)}{N} \left(\boldsymbol{\gamma}_{\ell+k}^{(m)}\right)^T.$$

Let

$$G_N^{(m)}(i,j) = \sum_{\ell=1}^{N-(m+1)} \gamma_\ell^{(m)}(i) H_{\ell,N}^{(m)}(j), \quad 1 \le i,j \le pq,$$

where $\gamma_\ell^{(m)}(i)$ and $H_{\ell,N}^{(m)}(j)$ are the i^{th} and the j^{th} coordinates of the vectors $\boldsymbol{\gamma}_{\ell,N}^{(m)}$ and $\mathbf{H}_{\ell,N}^{(m)}$, respectively. Next we write

$$\begin{aligned} E\left(G_N^{(m)}(i,j)\right)^2 &= E\left(\sum_{\ell=1}^{N-(m+1)} \gamma_\ell^{(m)}(i) H_{\ell,N}^{(m)}(j)\right)^2 \\ &= \sum\sum_{\substack{1\le r\le N-(m+1) \\ 1\le \ell\le N-(m+1)}} E\left(H_{\ell,N}^{(m)}(j)\gamma_\ell^{(m)}(i)\gamma_r^{(m)}(i)H_{r,N}^{(m)}(j)\right) \\ &= G_{1,N}^{(m)}(i,j) + G_{2,N}^{(m)}(i,j), \end{aligned}$$

where

$$G_{1,N}^{(m)}(i,j) = \sum\sum_{\substack{1\le r\le N-(m+1) \\ 1\le \ell\le N-(m+1) \\ |r-\ell|\le m}} E\left(H_{\ell,N}^{(m)}(j)\gamma_\ell^{(m)}(i)\gamma_r^{(m)}(i)H_{r,N}^{(m)}(j)\right),$$

and

$$G_{2,N}^{(m)}(i,j) = \sum\sum_{\substack{1\le r\le N-(m+1) \\ 1\le \ell\le N-(m+1) \\ |r-\ell|> m}} E\left(H_{\ell,N}^{(m)}(j)\gamma_\ell^{(m)}(i)\gamma_r^{(m)}(i)H_{r,N}^{(m)}(j)\right).$$

Notice that $\boldsymbol{\gamma}_\ell^{(m)}$ is independent of $\mathbf{H}_{\ell,N}^{(m)}$, $\mathbf{H}_{r,N}^{(m)}$ and $\boldsymbol{\gamma}_r^{(m)}$, if $r > m+\ell$. Hence

$$\begin{aligned} E\left(H_{\ell,N}^{(m)}(j)\gamma_\ell^{(m)}(i)\gamma_r^{(m)}(i)H_{r,N}^{(m)}(j)\right) &= \begin{cases} E\gamma_\ell^{(m)}(i)E\left(H_{\ell,N}^{(m)}(j)\gamma_r^{(m)}(i)H_{r,N}^{(m)}(j)\right) & r > m+\ell, \\ E\gamma_r^{(m)}(i)E\left(H_{\ell,N}^{(m)}(j)\gamma_\ell^{(m)}(i)H_{r,N}^{(m)}(j)\right) & \ell > m+r, \\ E\left(H_{\ell,N}^{(m)}(j)\gamma_\ell^{(m)}(i)\gamma_r^{(m)}(i)H_{r,N}^{(m)}(j)\right) & |\ell-r|\le m, \end{cases} \\ &= \begin{cases} 0 & |\ell-r| > m, \\ E\left(H_{\ell,N}^{(m)}(j)\gamma_\ell^{(m)}(i)\gamma_r^{(m)}(i)H_{r,N}^{(m)}(j)\right) & |\ell-r|\le m. \end{cases} \end{aligned}$$

Thus we have

$$EG_{2,N}^{(m)}(i,j) = 0.$$

Let M be an upper bound on $|K(t)|$. Using the fact that $\gamma_\ell^{(m)}$ is an m-dependent sequence, we now obtain the following:

$$E(H_{\ell,N}^{(m)}(j))^2 = \sum_{k=m+1}^{\min(N-\ell,B_N)} \sum_{v=m+1}^{\min(N-\ell,B_N)} \frac{K(k/B_N)}{N} \frac{K(v/B_N)}{N} E\left(\gamma_{\ell+k}^{(m)}(j)\gamma_{\ell+v}^{(m)}(j)\right) \tag{5.39}$$
$$\leq \frac{M^2}{N^2} \sum_{k=m+1}^{\min(N-\ell,B_N)} \sum_{v=m+1}^{\min(N-\ell,B_N)} E\left(\gamma_{\ell+k}^{(m)}(j)\gamma_{\ell+v}^{(m)}(j)\right)$$
$$\leq \frac{M^2}{N^2} B_N \sum_{r=-m}^{m} E\left|\gamma_0^{(m)}(j)\gamma_r^{(m)}(j)\right|$$
$$= O\left(\frac{B_N}{N^2}\right).$$

In the next step we will first use the Cauchy-Schwarz inequality, then the independence of $H_{\ell,N}^{(m)}(j)$ and $\gamma_\ell^{(m)}(i)$ and the independence of $H_{r,N}^{(m)}(j)$ and $\gamma_r^{(m)}(i)$ to get

$$\left|G_{2,N}^{(m)}(i,j)\right| \leq \sum\sum_{\substack{1\leq r\leq N-(m+1)\\ 1\leq \ell\leq N-(m+1)\\ |r-\ell|\leq m}} E\left|H_{\ell,N}^{(m)}(j)\gamma_\ell^{(m)}(i)\gamma_r^{(m)}(i)H_{r,N}^{(m)}(j)\right|$$
$$\leq \sum\sum_{\substack{1\leq r\leq N-(m+1)\\ 1\leq \ell\leq N-(m+1)\\ |r-\ell|\leq m}} \left(E\left(H_{\ell,N}^{(m)}(j)\gamma_\ell^{(m)}(i)\right)^2\right)^{1/2} \left(E\left(\gamma_r^{(m)}(i)H_{r,N}^{(m)}(j)\right)^2\right)^{1/2}$$
$$\leq \sum\sum_{\substack{1\leq r\leq N-(m+1)\\ 1\leq \ell\leq N-(m+1)\\ |r-\ell|\leq m}} \left(E\left(H_{\ell,N}^{(m)}(j)\right)^2\right)^{1/2} \left(E\left(\gamma_\ell^{(m)}(i)\right)^2\right)^{1/2} \left(E\left(\gamma_r^{(m)}(i)\right)^2\right)^{1/2}$$
$$\times \left(E\left(H_{r,N}^{(m)}(j)\right)^2\right)^{1/2}$$
$$\leq 2mNO\left(\frac{B_N^{1/2}}{N}\right)O(1)O(1)O\left(\frac{B_N^{1/2}}{N}\right)$$
$$= O\left(\frac{B_N}{N}\right)$$
$$= o(1),$$

where we also used (5.39) and Assumption 5.9. This completes the proof of Lemma 5.12. □

Let $\boldsymbol{i}^2 = -1$.

Lemma 5.13. *If Assumptions 5.1–5.4, 5.8 and 5.9 are satisfied, then for all $1 \leq j \leq pq$ we have*

$$\limsup_{N\to\infty} \limsup_{m\to\infty} \sup_{-\infty<t<\infty} E\left(\frac{1}{N^{1/2}} \sum_{k=1}^{N} (\gamma_k(j) - \gamma_k^{(m)}(j))e^{\boldsymbol{i}kt}\right)^2 = 0, \tag{5.40}$$

$$\limsup_{N\to\infty}\limsup_{m\to\infty}\sup_{-\infty<t<\infty} E\left(\frac{1}{N^{1/2}}\sum_{k=1}^{N}\gamma_k(j)e^{\boldsymbol{i}kt}\right)^2<\infty \tag{5.41}$$

and

$$\limsup_{N\to\infty}\limsup_{m\to\infty}\sup_{-\infty<t<\infty} E\left(\frac{1}{N^{1/2}}\sum_{k=1}^{N}\gamma_k^{(m)}(j)e^{\boldsymbol{i}kt}\right)^2<\infty. \tag{5.42}$$

Proof. First we note that

$$\begin{aligned}E\left(\sum_{k=1}^{N}(\gamma_k(j)-\gamma_k^{(m)}(j))e^{\boldsymbol{i}kt}\right)^2 &= \sum_{1\le k\le N} E((\gamma_k(j)-\gamma_k^{(m)}(j))e^{\boldsymbol{i}kt})^2\\ &\quad+2\sum_{1\le k<\ell\le N} E\left[(\gamma_k(j)-\gamma_k^{(m)}(j))(\gamma_\ell(j)-\gamma_\ell^{(m)}(j))\right]e^{\boldsymbol{i}(k+\ell)t}.\end{aligned}$$

It follows from (5.28) that there is a sequence $c_1(m)\to 0$ such that

$$\left|\sum_{1\le k\le N} E(\gamma_k(j)-\gamma_k^{(m)}(j))^2 e^{\boldsymbol{i}2kt}\right|\le Nc_1(m).$$

Next we write

$$\begin{aligned}&\sum_{1\le k<\ell\le N} E\left[(\gamma_k(j)-\gamma_k^{(m)}(j))\gamma_\ell(j)\right]e^{\boldsymbol{i}(k+\ell)t}\\ &\qquad=\sum_{1\le k<\ell\le N} E\left[(\gamma_k(j)-\gamma_k^{(m)}(j))(\gamma_\ell(j)-\gamma_\ell^{(\ell-k)}(j))\right]e^{\boldsymbol{i}(k+\ell)t},\end{aligned}$$

since $(\boldsymbol{\gamma}_k,\boldsymbol{\gamma}_k^{(m)})$ and $\boldsymbol{\gamma}_\ell^{(\ell-k)}$ are independent. Using the Cauchy-Schwarz inequality first, then (5.28) again we get that

$$\begin{aligned}&\sum_{1\le k<\ell\le N}\left|E\left[(\gamma_k(j)-\gamma_k^{(m)}(j))(\gamma_\ell(j)-\gamma_\ell^{(\ell-k)}(j))\right]e^{\boldsymbol{i}(k+\ell)t}\right|\\ &\qquad\le\sum_{1\le k<\ell\le N}\left[E(\gamma_k(j)-\gamma_k^{(m)}(j))^2\right]^{1/2}\left[E(\gamma_\ell(j)-\gamma_\ell^{(\ell-k)}(j))^2\right]^{1/2}\\ &\qquad\le N\left[E(\gamma_1(j)-\gamma_1^{(m)}(j))^2\right]^{1/2}\sum_{1\le k<\infty}\left[E(\gamma_1(j)-\gamma_1^{(k)}(j))^2\right]^{1/2}\\ &\qquad=Nc_2(m)\end{aligned}$$

with some sequence $c_2(m)\to 0$. Similar arguments show that

$$\sum_{1\le k<\ell\le N}\left|E\left[(\gamma_k(j)-\gamma_k^{(m)}(j))\gamma_\ell^{(m)}(j)\right]e^{\boldsymbol{i}(k+\ell)t}\right|=Nc_3(m)$$

with some sequence $c_3(m)\to 0$, completing the proof of (5.40).

Similarly to the proof of (5.40), we write

$$\begin{aligned}
E\left(\sum_{k=1}^{N}\gamma_k(j)e^{\boldsymbol{i}kt}\right)^2 &= \sum_{k=1}^{N}\sum_{\ell=1}^{N}E\gamma_k(j)\gamma_\ell(j)e^{\boldsymbol{i}(k+\ell)t}\\
&= \sum_{k=1}^{N}E\gamma_k^2(j)e^{2\boldsymbol{i}kt}+2\sum_{1\le k<\ell\le N}E\gamma_k(j)\gamma_\ell(j)e^{\boldsymbol{i}(k+\ell)t}\\
&= E\gamma_1^2(j)\sum_{k=1}^{N}e^{2\boldsymbol{i}kt}+2\sum_{1\le k<\ell\le N}E\gamma_k(j)(\gamma_\ell(j)-\gamma_\ell^{(\ell-k)}(j))e^{\boldsymbol{i}(k+\ell)t},
\end{aligned}$$

since by the independence of $\gamma_k(j)$ and $\gamma_\ell^{(\ell-k)}(j)$ we have that $E\gamma_k(j)\gamma_\ell^{(\ell-k)}(j)=0$. Using the Cauchy-Schwarz inequality with (5.28) we get that

$$\left|\sum_{1\le k<\ell\le N}E\gamma_k(j)(\gamma_\ell(j)-\gamma_\ell^{(\ell-k)}(j))e^{\boldsymbol{i}(k+\ell)t}\right|\le cN$$

with some constant c, completing the proof of (5.41). The same arguments can be used to prove (5.42). □

Following Taniguchi and Kakizawa (2000) we define $\mathbf{S}_N(t)=\sum_{k=1}^{N}\boldsymbol{\gamma}_{k,N}e^{\boldsymbol{i}kt}$ and $\mathbf{S}_N^{(m)}(t)=\sum_{k=1}^{N}\boldsymbol{\gamma}_{k,N}^{(m)}e^{\boldsymbol{i}kt}$. Let $\mathbf{S}_N^*(t)$ be the conjugate transpose of $\mathbf{S}_N(t)$ and introduce

$$\begin{aligned}
\mathbf{I}_N(t) &= \frac{1}{N}\mathbf{S}_N(t)\mathbf{S}_N^*(t)\\
&= \frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\gamma}_k e^{\boldsymbol{i}kt}\sum_{\ell=1}^{N}\boldsymbol{\gamma}_\ell^T e^{-\boldsymbol{i}\ell t}\\
&= \frac{1}{N}\sum_{\ell=1}^{N}\sum_{k=1}^{N}e^{\boldsymbol{i}t(k-\ell)}\boldsymbol{\gamma}_k\boldsymbol{\gamma}_\ell^T\\
&= \sum_{k=1-N}^{N-1}e^{-\boldsymbol{i}tk}\frac{1}{N}\sum_{\ell=\max(1,1-k)}^{\min(N,N-k)}\boldsymbol{\gamma}_k\boldsymbol{\gamma}_{\ell+k}^T\\
&= \sum_{k=1-N}^{N-1}e^{-\boldsymbol{i}tk}\boldsymbol{\phi}_{k,N}.
\end{aligned}$$

Similarly we define

$$\mathbf{I}_N^{(m)}(t)=\frac{1}{N}\mathbf{S}_N^{(m)}(t)\left(\mathbf{S}_N^{(m)}(t)\right)^*=\sum_{k=1-N}^{N-1}e^{-\boldsymbol{i}tk}\boldsymbol{\phi}_{k,N}^{(m)}.$$

Lemma 5.14. *If Assumptions 5.1–5.4, 5.8 and 5.9 are satisfied, then we have*

$$\limsup_{N\to\infty}\limsup_{m\to\infty}\sup_{-\infty<t<\infty}E\left|\mathbf{I}_N(t)-\mathbf{I}_N^{(m)}(t)\right|=0.$$

Proof. By the triangle inequality we have

$$\begin{aligned}\left|\mathbf{I}_N(t)-\mathbf{I}_N^{(m)}(t)\right| &= \left|\frac{1}{N}\mathbf{S}_N(t)\mathbf{S}_N^*(t)-\frac{1}{N}\mathbf{S}_N^{(m)}(t)\left(\mathbf{S}_N^{(m)}(t)\right)^*\right| \\ &\le \frac{1}{N}\left|\mathbf{S}_N(t)(\mathbf{S}_N^*(t)-(\mathbf{S}_N^{(m)}(t))^*)\right| \\ &\quad + \frac{1}{N}\left|(\mathbf{S}_N(t)-\mathbf{S}_N^{(m)}(t))(\mathbf{S}_N^{(m)}(t))^*\right|.\end{aligned}$$

Now the result follows from Lemma 5.13 via the Cauchy-Schwartz inequality. □

Proof of Theorem 5.2. Define the Fourier transform, $\hat{K}(u)$, of the kernel K as $\hat{K}(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} K(s)e^{-\boldsymbol{i}su}\,ds$. Since K and $\hat{K}$ are in L^1 and both are Lipschitz functions, the inversion formula gives $K(s)=\int_{-\infty}^{\infty}\hat{K}(u)e^{\boldsymbol{i}su}\,du$. From the relationship between K and $\hat{K}$ and from the fact that K is supported on the interval $[-1,1]$, we obtain:

$$\begin{aligned}\tilde{\mathbf{\Sigma}}_N &= \sum_{k=-B_N}^{B_N} K(k/B_N)\boldsymbol{\phi}_{k,N} \\ &= \sum_{k=1-N}^{N-1} K(k/B_N)\boldsymbol{\phi}_{k,N} \\ &= \sum_{k=1-N}^{N-1}\left(\int_{-\infty}^{\infty}\hat{K}(u)e^{\boldsymbol{i}(k/B_N)u}\,du\right)\boldsymbol{\phi}_{k,N} \\ &= \int_{-\infty}^{\infty}\hat{K}(u)\sum_{k=1-N}^{N-1}e^{-\boldsymbol{i}(-u/B_N)k}\boldsymbol{\phi}_{k,N}\,du \\ &= \int_{-\infty}^{\infty}\hat{K}(u)\mathbf{I}_N(-u/B_N)\,du.\end{aligned}$$

Similarly,

$$\tilde{\mathbf{\Sigma}}_N^{(m)} = \int_{-\infty}^{\infty}\hat{K}(u)\mathbf{I}_N^{(m)}(-u/B_N)\,du.$$

Hence we have

$$\begin{aligned}E\left|\tilde{\mathbf{\Sigma}}_N-\tilde{\mathbf{\Sigma}}_N^{(m)}\right| &= E\left|\int_{-\infty}^{\infty}\hat{K}(u)\left(\mathbf{I}_N(u/B_N)-\mathbf{I}_N^{(m)}(u/B_N)\right)du\right| \\ &\le \int_{-\infty}^{\infty}\left|\hat{K}(u)\right|E\left|\left(\mathbf{I}_N(u/B_N)-\mathbf{I}_N^{(m)}(u/B_N)\right)\right|du \\ &\le \sup_{-\infty<t<\infty}\left|\left|\mathbf{I}_N(t)-\mathbf{I}_N^{(m)}(t)\right|\right|_1\int_{-\infty}^{\infty}\left|\hat{K}(u)\right|du.\end{aligned}$$

Applying Lemma 5.14 we conclude that

$$\left|\tilde{\mathbf{\Sigma}}_N-\tilde{\mathbf{\Sigma}}_N^{(m)}\right|\xrightarrow{P}0,$$

as $\min(N,m)\to\infty$. On the other hand, by Lemma 5.12, for every fixed m

$$\tilde{\mathbf{\Sigma}}_N^{(m)} \xrightarrow{P} \mathbf{\Sigma}^{(m)}.$$

Since

$$\mathbf{\Sigma}^{(m)} \to \mathbf{\Sigma},$$

as $m\to\infty$, the proof of the theorem is complete. $\square$

Proof of Theorem 5.3. It follows from the definition of $\hat{\epsilon}_\ell$, (5.4) and the orthonormality of $\{w_j, 1\le j<\infty\}$ that

$$\langle \hat{\epsilon}_\ell, w_i\rangle = \langle \epsilon_\ell, w_i\rangle + \langle X_\ell, u_i\rangle + \langle \nu_\ell, w_i\rangle,$$

where

$$\nu_\ell(t) = \sum_{i=1}^{q}\sum_{j=1}^{p}\psi_{i,j}w_i(t)\langle X_\ell, v_j\rangle - \sum_{i=1}^{q}\sum_{j=1}^{p}\hat{\psi}_{i,j}\hat{w}_{i,N}(t)\langle X_\ell, \hat{v}_{j,N}\rangle.$$

Following the proof of Theorem 5.2 one can show that the estimates in (5.20) and (5.21) yield

$$\left|\breve{\Sigma}_N(i,j,i',j') - \hat{d}_{i,N}\hat{c}_{j,N}\hat{d}_{i',N}\hat{c}_{j',N}\Sigma_N^*(i,j,i',j')\right| = o_P(1), \tag{5.43}$$

where

$$\breve{\Sigma}_N(i,j,i',j') = \sum_{k=-(N-1)}^{N-1} K(k/B_N)\hat{\phi}_{k,N}(i,j,i',j')$$

and

$$\Sigma_N^*(i,j,i',j') = \sum_{k=-(N-1)}^{N-1} K(k/B_N)\phi_{k,N}^*(i,j,i',j')$$

with

$$\hat{\phi}_{k,N}(i,j,i',j') = \frac{1}{N}\sum_{\ell=\max(1,1-k)}^{\min(N,N-k)}\hat{\gamma}_\ell(i,j)\hat{\gamma}_{\ell+k}(i',j'),$$

$$\phi_{k,N}^*(i,j,i',j') = \frac{1}{N}\sum_{\ell=\max(1,1-k)}^{\min(N,N-k)}\gamma_\ell^*(i,j)\gamma_{\ell+k}^*(i',j'),$$

and

$$\gamma_\ell^*(i,j) = \langle X_\ell, v_j\rangle\langle \hat{\epsilon}_\ell, w_i\rangle.$$

Since

$$\langle X_\ell, v_j\rangle\langle \hat{\epsilon}_\ell, w_i\rangle = \gamma_\ell(i,j) + \langle X_\ell, v_j\rangle\langle \nu_\ell, w_i\rangle,$$

(5.20), (5.21) and Lemma 5.10 imply that

$$\left|\tilde{\Sigma}_N - \Sigma_N^*\right| = o_P(1). \tag{5.44}$$

□

We have seen in Theorem 5.2 that $\left|\tilde{\mathbf{\Sigma}}_N - \mathbf{\Sigma}\right| = o_P\,(1)$. In (5.43) and (5.44) we have seen that $\left|\breve{\mathbf{\Sigma}}_N - \boldsymbol{\zeta}_N \mathbf{\Sigma}_N^* \boldsymbol{\zeta}_N\right| = o_P\,(1)$ and $\left|\tilde{\Sigma}_N - \Sigma_N^*\right| = o_P(1)$. Therefore, $\left|\hat{\mathbf{\Sigma}}_N - \mathbf{\Sigma}\right| = o_P\,(1)$, completing the proof.

5.7 Bibliography

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley and Sons.

Aue, A., S. Hörmann, L. Horváth, M. Hušková, and J. Steinebach (2011). Sequential Testing for the Stability of Portfolio Betas. *Econometric Theory 00*, 000–000.

Aue, A., S. Hörmann, L. Horváth, and M. Reimherr (2009). Break Detection in the Covariance Structure of Multivariate Time Series Models. *The Annals of Statistics 37*, 4046–4087.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer.

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory.* Holt.

Cai, T. and P. Hall (2006). Prediction in Functional Linear Regression. *The Annals of Statistics 34*, 2159–2179.

Cardot, H., F. Ferraty, A. Mas, and P. Sarda (2003). Testing Hypothesis in the Functional Linear Model. *Scandinavian Journal of Statistics 30*, 241–255.

Chiou, J.-M., H.-G. Müller, and J.-L. Wang (2004). Functional Response Models. *Statistica Sinica 14*, 675–693.

Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer.

Gohberg, I., S. Golberg, and M. A. Kaashoek (1990). *Classes of Linear Operators*, Volume 49 of *Operator Theory: Advances and Applications.* Birkhaüser.

Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series.* Wiley.

Hörmann, S., L. Horváth, and R. Reeder (2010). A Functional Version of the ARCH Model. Technical report, University of Utah.

Hörmann, S. and P. Kokoszka (2010). Weakly Dependent Functional Data. *The Annals of Statistics 38*, 1845–1884.

Horváth, L., M. Hušková, and P. Kokoszka (2010). Testing the Stability of the Functional Autoregressive Process. *Journal of Multivariate Analysis 101*, 352–367.

Ibragimov, I. (1962). Some Limit Theorems for Stationary Processes. *Theory of Probability and its Applications 7*, 349–382.

Jansson, M. (2002). Consistent Covariance Matrix Estimation for Linear Processes. *Econometric Theory 18*, 1449–1459.

Kargin, V. and A. Onatski (2008). Curve Forecasting by Functional Autoregression. *Journal of Multivariate Analysis 99*, 2508–2526.

Kiefer, J. (1959). K-sample Analogues of the Kolmogorov-Smirnov and Cramér-v.Mises Tests. *Ann. Math. Statist. 30*, 420–447.

Liu, W. and W. B. Wu (2010). Asymptotics of Spectral Density Estimates. *Econometric Theory 26*, 1218–1245.

Móricz, F. (1976). Moment Inequalities and the Strong Law of Large Numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 35*, 299–314.

Parzen, E. (1961). Mathematical Considerations in the Estimation of Spectra. *Technometrics 3*, 167–190.

Parzen, E. (1967). *Time Series Analysis Papers.* Holden-Day.

Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Academic Press.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Birkhäuser.

Taniguchi, M. and Y. Kakizawa (2000). *Asymptotic Theory of Statistical Inference for Time Series.* Springer.

CHAPTER 6

A FUNCTIONAL VERSION OF THE ARCH MODEL[4]

Improvements in data acquisition and processing techniques have lead to an almost continuous flow of information for financial data. High resolution tick data are available and can be quite conveniently described by a continuous time process. It is therefore natural to ask for possible extensions of financial time series models to a functional setup. In this chapter we propose a functional version of the popular ARCH model. We will establish conditions for the existence of a strictly stationary solution, derive weak dependence and moment conditions, show consistency of the estimators and perform a small empirical study demonstrating how our model matches with real data.

6.1 Introduction

To date not many functional time series models exist to describe sequences of dependent observations. Arguably the most popular is the ARH(1), the autoregressive Hilbertian process of order 1. It is a natural extension of the scalar and vector valued AR(1) process (Brockwell and Davis, 1991). Due to the fact that the ARH(1) model is mathematically and statistically quite flexible and well established, it is used in practice for modeling and prediction of continuous-time random experiments. We refer to Bosq (2000) for a detailed treatment of moving averages, autoregressive and general linear time series sequences. Despite the prominent presence in time series analysis it is clear that the applicability of moving average and autoregressive processes is limited. To describe nonlinear models in the scalar and vector cases, a number of different approaches have been introduced in the last decades. One of the most popular ones in econometrics is the ARCH model of Engle (1982) and the more general GARCH model of Bollerslev (1990) which have had an enormous impact on the modeling of financial data. For surveys on volatility models we refer to Silvennoinen and Teräsvirta (2009). GARCH-type models are designed for the

[4]The content of this chapter is based on joint research with Siegfried Hörmann and Lajos Horváth. It has been submitted to Econometric Theory.

analysis of daily, weekly or more general long-term period returns. Improvements in data acquisition and processing techniques have lead to an almost continuous flow of information for financial data with online investment decisions. High resolution tick data are available and can be quite conveniently described as functions. It is therefore natural to ask for possible extensions of these financial time series models to a functional setup. The idea is that instead of a scalar return sequence $\{y_k, 1 \le k \le T\}$ we have a functional time series $\{y_k(t), 1 \le k \le T, 0 \le t \le S\}$, where $y_k(t)$ are intraday (log-)returns on day k at time t. In other words if $\{P_k(t), 1 \le k \le T, 0 \le t \le S\}$ is the underlying price process, then $y_k(t) = \log P_k(t) - \log P_k(t-h)$ for the desired time lag h, where we will typically set $h = 5\text{min}$. By rescaling we can always assume that $S = 1$ and then the interval $[0, 1]$ represents one trading day.

We notice that a daily segmentation of the data is natural and preferable to only one continuous time process $\{y(s), 0 \le s \le T\}$, say, for all T days of our sample (Harrison et al., 1984; Barndorff-Nielsen and Shephard, 2004; Zhang et al., 2005; Barndorff-Nielsen et al., 2008; Jacod et al., 2009). Due to the time laps between trading days (implying, e.g., that opening and closing prices do not necessarily coincide) one continuous time model might not be suitable for a longer period. Intraday volatilities of the euro-dollar rates investigated by Cyree et al. (2004) empirically can be considered as daily curves. Similarly, Gau (2005) studied the shape of the intraday volatility curves of the Taipei FX market. Angelidis and Degiannakis (2008) compared predictions based on intraday and interday data. Elezović (2009) modeled bid and ask prices as continuous functions. The spot exchange rates in Fatum and Pedersen (2009) can be considered as functional observations as well. Evans and Speight (2010) uses 5-min returns for Euro-Dollar, Euro-Sterling and Euro-Yen exchange rates.

In this chapter we propose a functional ARCH model. Usually time series are defined by stochastic recurrence equations establishing the relationship between past and future observations. The question preceding any further analysis is whether such an equation has a (stationary) solution. For the scalar ARCH necessary and sufficient conditions have been derived by Nelson (1990). Interestingly, these results cannot be transferred directly to multivariate extensions (Silvennoinen and Teräsvirta, 2009). Due to the complicated dynamics of multivariate ARCH/GARCH type models (MGARCH), finding the necessary and sufficient conditions for the existence of stationary solutions to the defining equations is a difficult problem. Also the characterization of the existence of the moments in GARCH(p, q) equations is given by very involved formulas (Ling and McAleer, 2002). It is therefore not surprising that in a functional setup, i.e. when dealing with intrinsically infinite dimensional objects, some balancing between generality and mathematical feasibility of the model is required.

In Section 6.2 we propose a model for which we provide conditions for the existence of a unique stationary solution. These conditions are not too far from being optimal. We will also study the dependence structure of the model, which is useful in many applications, e.g. in estimation which will be treated in Section 6.3. We also provide an example illustrating that the proposed functional ARCH model is able to capture typical characteristics of high frequency returns, see Section 6.4.

In this chapter we use the following notation. Let F denote a generic function space. Throughout this consists of real valued functions with domain $[0,1]$. In many applications F will be equal to $\mathcal{H}=L^2([0,1])$, the Hilbert space of square integrable functions with norm $\|x\|_{\mathcal{H}} = \big(\int_0^1 x^2(s)\,ds\big)^{1/2}$ which is generated by the inner product $\langle x,y\rangle = \int_0^1 x(s)y(s)\,ds$ for $x,y \in \mathcal{H}$. Another important example is $F = \mathcal{C}[0,1]$. This is the space of continuous functions on $[0,1]$ equipped with the sup-norm $\|x\|_\infty = \sup_{t\in[0,1]} |x(t)|$. By F^+ we denote the set of non-negative functions in F. To further lighten notation we shall often write x when we mean $\{x(t),\, t\in[0,1]\}$, or β for integral kernels $\{\beta(t,s),\, 0\le t\le 1,\, 0\le s\le 1\}$ as well as for the corresponding operators. If $x,y\in F$ then xy stands for pointwise multiplications, i.e. $xy = \{x(s)y(s),\, s\in[0,1]\}$. Since integrals will always be taken over the unit interval we shall henceforth simply write $\int x(t)\,dt$. A random function X with values in $\mathcal{H}$ is said to be in $L^p_{\mathcal{H}}$ if $\nu_p(X) = \big(E\|X\|^p_{\mathcal{H}}\big)^{1/p} < \infty$.

6.2 The functional ARCH model

We start with the following general definition.

Definition 6.1. *Let $\{\varepsilon_k\}$ be a sequence of independent and identically distributed random functions in F. Further let $\beta : F^+ \to F^+$ be a non-negative operator and let $\delta \in F^+$. Then an F-valued process $\{y_k(s),\, k\in\mathbb{Z},\, s\in[0,1]\}$ is called a functional ARCH(1) process in F if the following holds:*

$$y_k = \varepsilon_k \sigma_k \tag{6.1}$$

and

$$\sigma_k^2 = \delta + \beta(y_{k-1}^2). \tag{6.2}$$

The assumption for the existence of processes satisfying (6.1) and (6.2) depends on the choice of F. So next we specify F and put some restrictions on the operator β. Our first result gives a sufficient condition for the existence of a strictly stationary solution when $F = \mathcal{H}$. We will assume that β is a (bounded) kernel operator defined by

$$\beta(x)(t) = \int \beta(t,s)x(s)\,ds, \quad x \in \mathcal{H}. \tag{6.3}$$

Boundedness is e.g. guaranteed by finiteness of the Hilbert-Schmidt norm:

$$\|\beta\|_{\mathcal{S}}^2 = \int\int \beta^2(t,s)\,ds\,dt < \infty. \tag{6.4}$$

Theorem 6.1. *Let $\{y_k\}$ be the process given in Definition 6.1 with $F = \mathcal{H}$ and β given in* (6.3)*, such that the operator β is bounded. Define $K(\varepsilon_1^2) = \big(\int\int \beta^2(t,s)\varepsilon_1^4(s)\,ds\,dt\big)^{1/2}$. If there is some $\alpha > 0$ such that $E\big\{K(\varepsilon_1^2)\big\}^{\alpha} < 1$, then* (6.1) *and* (6.2) *have a unique strictly stationary solution in $\mathcal{H}$. Furthermore, σ_k^2 is of the form*

$$\sigma_k^2 = g(\varepsilon_{k-1}, \varepsilon_{k-2}, \ldots), \tag{6.5}$$

with some measurable functional $g : \mathcal{H}^{\mathbb{N}} \to \mathcal{H}$.

It follows that $\{\sigma_k\}$ and $\{y_k\}$ are not just strictly stationary but also ergodic (Stout, 1974). Let $\mathcal{F}_k$ be the σ-algebra generated by the sequence $\{\varepsilon_i,\, i \le k\}$. If (6.1) and (6.2) have a stationary solution and if we assume that $E\varepsilon_k(t) = 0$, $E\varepsilon^2(t) < \infty$ and $E\sigma_k^2(t) < \infty$ for all $t \in [0,1]$, then due to (6.5) it is easy to see that

$$\operatorname{corr}\big\{y_k(t), y_k(s)|\mathcal{F}_{k-1}\big\} = \operatorname{corr}\big\{\varepsilon_k(t), \varepsilon_k(s)\big\}.$$

Since by our assumption $\{\varepsilon_k,\, k \in \mathbb{Z}\}$ is stationary, the conditional correlation is independent of k and can be fully described by the covariance kernel $C_\varepsilon(t,s) = \operatorname{Cov}(\varepsilon(t), \varepsilon(s))$. However, we have $\operatorname{Cov}\big\{y_k(t), y_k(s)|\mathcal{F}_{k-1}\big\} = \sigma_k(t)\sigma_k(s)C_\varepsilon(s,t)$. This is in accordance with the *constant conditional correlation* (CCC) multivariate GARCH models of Bollerslev (1990) and Jeantheau (1998).

Our next result shows that σ_k^2 of (6.5) can be geometrically approximated with m-dependent variables, which establishes weak dependence of the processes (6.1) and (6.2).

Theorem 6.2. *Assume that the conditions of Theorem 6.1 hold. Let $\{\varepsilon_k'\}$ be an independent copy of $\{\varepsilon_k\}$ and define $\sigma_{km}^2 = g(\varepsilon_{k-1}, \varepsilon_{k-2}, \ldots, \varepsilon_{k-m}, \varepsilon_{k-m-1}', \varepsilon_{k-m-2}', \ldots)$. Then*

$$E\{\|\sigma_k^2 - \sigma_{km}^2\|_{\mathcal{H}}\}^{\alpha} \le cr^m, \tag{6.6}$$

with some $0 < r = r(\alpha) < 1$ and $c = c(\alpha) < \infty$.

To better understand the idea behind our result we remark the following. Assume that we redefine

$$\sigma_{km}^2 = g(\varepsilon_{k-1}, \varepsilon_{k-2}, \ldots, \varepsilon_{k-m}, \varepsilon_{k-m-1,k-m}^{(k)}, \varepsilon_{k-m-2,k-m}^{(k)}, \ldots),$$

where $\{\varepsilon_{\ell,i}^{(k)}, \ell, i, k \in \mathbb{Z}\}$ are independent copies of $\{\varepsilon_\ell, \ell \in \mathbb{Z}\}$. In other words, every σ_k^2 gets its "individual" copy of $\{\varepsilon_{\ell,i}^{(k)}\}$ to define the approximations. It can be easily seen that then for any fixed $m \geq 1$, $\{\sigma_{km}^2,\ k \in \mathbb{Z}\}$ form m-dependent sequences, while the value on the left hand side in inequality (6.6) doesn't change. As we have shown in our recent papers, approximations like (6.6) are particularly useful in studying large sample properties of functional data (Aue et al., 2011; Hörmann and Kokoszka, 2010). We use (6.6) to provide conditions for the existence of moments of the stationary solutions of (6.1) and (6.2). It also follows immediately from (6.6), that if (6.1) and (6.2) are solved starting with some initial values y_0^* and σ_0^*, then the effect of the initial values dies out exponentially fast.

In a finite dimensional vector space all norms are equivalent. This is no longer true in the functional (infinite dimensional) setup and whether a solution of (6.1) and (6.2) exists depends on the choice of space and norm of the state space. Depending on the application, it might be more convenient to work in a different space. We give here the analogue of Theorems 6.1 and 6.2 for a functional ARCH process in $\mathcal{C}[0,1]$.

Theorem 6.3. *Let $\{y_k\}$ be the process given in Definition 6.1 with $F = \mathcal{C}[0,1]$ and define $H(\varepsilon_1^2) = \sup_{0 \leq t \leq 1} \int \beta(t,s)\varepsilon_1^2(s)\,ds$. If there is some $\alpha > 0$ such that $E\{H(\varepsilon_1^2)\}^\alpha < 1$, then* (6.1) *and* (6.2) *have a unique strictly stationary solution in $\mathcal{C}[0,1]$. Furthermore, σ_k^2 can be represented as in* (6.5). *In addition the proposition of Theorem 6.2 holds, with* (6.6) *replaced by*

$$E\{\|\sigma_k^2 - \sigma_{km}^2\|_\infty\}^\alpha \leq cr^m.$$

We continue with some immediate consequences of our theorems. We start with conditions for the existence of the moments of the stationary solution of (6.1) and (6.2).

Proposition 6.1. *Assume that the conditions of Theorem 6.1 hold. Then*

$$E\{\|\sigma_0^2\|_{\mathcal{H}}\}^\alpha < \infty. \tag{6.7}$$

If

$$E\{\|\sigma_0\|_{\mathcal{H}}\}^\alpha < \infty \tag{6.8}$$

and

$$E\{\|\varepsilon_0\|_\infty\}^\alpha < \infty, \tag{6.9}$$

then

$$E\{\|y_0\|_{\mathcal{H}}\}^{\alpha} < \infty. \tag{6.10}$$

Proposition 6.2. *Assume that the conditions of Theorem 6.3 hold. Then the analogue of Proposition 6.1 holds, with* $\|\cdot\|_{\mathcal{H}}$ *in* (6.7)–(6.10) *replaced by* $\|\cdot\|_{\infty}$.

We would like to point out that it is not assumed that the innovations ε_k have finite variance. We only need that ε_k have some moment of order $\alpha > 0$, where $\alpha > 0$ can be as small as we wish. Hence our model allows for innovations as well as observations with heavy tails.

According to Propositions 6.1 and 6.2, if the innovation ε_0 has enough moments, then so does σ_0^2 and y_0. The next result shows a connection between the moduli of continuity of ε_0 and y_0. Let

$$\omega(x, h) = \sup_{0\le t\le 1-h} \sup_{0\le s\le h} |x(t+s) - x(t)|$$

denote the modulus of continuity of a function $x(t)$.

Proposition 6.3. *We assume that the conditions of Theorem 6.3 are satisfied with* $\alpha = p > 0$. *If* $E\{\|\varepsilon_0\|_{\infty}\}^p < \infty$ *and* $\lim_{h\to 0} E\{\omega(\varepsilon_0, h)\}^p = 0$, *then* $\lim_{h\to 0} E\{\omega(y_0, h)\}^p = 0$.

According to Theorems 6.2 and 6.3, the stationary solution of (6.1) and (6.2) can be approximated with stationary, weakly dependent sequences with values in $\mathcal{H}$ and in $\mathcal{C}[0,1]$, respectively. We provide two further results which establish the weak dependence structure of $\{y_k\}$.

Proposition 6.4. *We assume that the conditions of Theorem 6.1 are satisfied with* $\alpha = \frac{p}{2}$ *and*

$$E\{\|\varepsilon_0\|_{\infty}\}^p < \infty. \tag{6.11}$$

Then

$$E\{\|y_k - y_{km}\|_{\mathcal{H}}\}^p \le c\gamma^m, \quad -\infty < k < \infty, m \ge 1, \tag{6.12}$$

with some $0 < c < \infty$ *and* $0 < \gamma < 1$, *where* $y_{km} = \varepsilon_k \sigma_{km}$.

It follows from the definitions that the distribution of the $y_k - y_{km}$ does not depend on k. Hence the expected value in (6.12) does not depend on k. A similar result holds in $F = \mathcal{C}[0,1]$ under the sup-norm.

Proposition 6.5. *We assume that the conditions of Theorem 6.3 are satisfied with* $\alpha = \frac{p}{2}$ *and that* (6.11) *holds. Then*

$$\sup_{0\le t\le 1} E\big|y_k - y_{km}|^p \le c\gamma^m, \quad -\infty < k < \infty, m \ge 1, \tag{6.13}$$

with some $0 < c < \infty$ *and* $0 < \gamma < 1$*, where* $y_{km} = \varepsilon_k \sigma_{km}$*.*

As in case of Proposition 6.4, the expected value in (6.13) does not depend on k.

6.3 Estimation

In this section we propose estimators for the function δ and the operator β in model (6.1)–(6.2) which are not known in practice. The procedure is developed for the important case where $F = \mathcal{H}$ and β is given as in (6.3). We show that our problem is related to the estimation of the autocorrelation operator in the ARH(1) model which has been intensively studied in Bosq (2000). However, the theory developed in Bosq (2000) is not directly applicable as it requires independent innovations in the ARH(1) process, whereas, as we will see below, we can only assume weak white noise (in Hilbert space sense).

We will impose the following

Assumption 6.1. ***(a)*** $E\varepsilon_0^2(t) = 1$ *for any* $t \in [0, 1]$*.*

(b) *The assumptions of Theorem 6.2 hold with* $\alpha = 2$*.*

Assumption 6.1 (a) is needed to guarantee the identifiability of the model. Part (b) of the assumption guarantees the existence of a stationary solution of the model (6.1)–(6.2) with moments of order 4. It is necessary to make the moment based estimator proposed below working. An immediate consequence of Assumption 6.1 is that (6.4) holds, i.e. β is a Hilbert Schmidt operator.

We let m_2 denote the mean function of the y_k^2 and introduce

$$\nu_k = y_k^2 - \sigma_k^2 = \{(\varepsilon_k^2(s) - 1)\sigma_k^2(s),\, s \in [0, 1]\}.$$

Then by adding ν_k on both sides of (6.2) we obtain

$$y_k^2 = \delta + \beta(y_{k-1}^2) + \nu_k.$$

Since β is a linear operator we obtain after subtracting m_2 on both sides of the above equation

$$y_k^2 - m_2 = \delta - m_2 + \beta(m_2) + \beta(y_{k-1}^2 - m_2) + \nu_k. \tag{6.14}$$

It can be easily seen that under Assumption 6.1 $E\nu_k = 0$ (where 0 stands for the zero function). Notice also that the expectation commutes with bounded operators, and hence that $E(\beta(y_k^2 - m_2)) = \beta(E(y_k^2 - m_2)) = 0$. Consequently, taking expectations on both sides of (6.14) yields that

$$\delta - m_2 + \beta(m_2) = 0. \tag{6.15}$$

Thus, (6.14) can be rewritten in the form

$$Z_k = \beta(Z_{k-1}) + \nu_k \quad \text{with} \quad Z_k = y_k^2 - m_2. \tag{6.16}$$

Model (6.16) is the autoregressive Hilbertian model of order 1, short ARH(1). For estimating the autocorrelation operator β we may use the estimator proposed in chapter 8 of Bosq (2000). We need to be aware, however, that the theory in Bosq (2000) has been developed for ARH processes with strong white noise innovations, i.e. independent innovations $\{\nu_k\}$. In our setup the $\{\nu_k\}$ form only a *weak white noise* sequence, i.e. for any $n \neq m$ we have

$$E\|\nu_n\|_{\mathcal{H}}^2 < \infty \quad \text{and} \quad E\langle \nu_n, x\rangle\langle \nu_m, y\rangle = 0 \quad \forall x, y \in \mathcal{H},$$

and the covariance operator of ν_n is independent of n. Thus the theory in Bosq (2000) cannot be directly applied. We will study the estimation of β in Section 6.3.1.

Once β is estimated by some $\hat{\beta}$ say, we obtain an estimator for δ via equation (6.15):

$$\hat{\delta} = \hat{m}_2 - \hat{\beta}(\hat{m}_2), \tag{6.17}$$

where we use

$$\hat{m}_2 = \frac{1}{N}\sum_{k=1}^{N} y_k^2. \tag{6.18}$$

Let $\|\beta\|_{\mathcal{L}} = \sup_{x\in\mathcal{H}}\{\|\beta(x)\|_{\mathcal{H}} : \|x\| \leq 1\}$ be the operator norm of β. Recall that $\|\beta\|_{\mathcal{L}} \leq \|\beta\|_{\mathcal{S}}$. The following Lemma shows that consistency of $\hat{\beta}$ implies consistency of $\hat{\delta}$.

Lemma 6.1. *Let Assumption 6.1 hold. Let $\hat{\delta} = \hat{\delta}_N$ be given as in* (6.17). *Then*

$$\|\hat{\delta}_N - \delta\|_{\mathcal{H}} = O_P(1) \times \left(N^{-1/2} + \|\hat{\beta}_N - \beta\|_{\mathcal{L}}\right).$$

Proof. We have

$$\begin{aligned}\|\hat{\delta}_N - \delta\|_{\mathcal{H}} &\le \|\hat{m}_2 - m_2\|_{\mathcal{H}} + \|\hat{\beta}(\hat{m}_2) - \beta(\hat{m}_2)\|_{\mathcal{H}} + \|\beta(\hat{m}_2) - \beta(m_2)\|_{\mathcal{H}} \\ &\le \|\hat{m}_2 - m_2\|_{\mathcal{H}} + \|\hat{\beta} - \beta\|_{\mathcal{L}}\|\hat{m}_2\|_{\mathcal{H}} + \|\beta\|_{\mathcal{L}}\|\hat{m}_2 - m_2\|_{\mathcal{H}}.\end{aligned}$$

The result follows once we can show that $\|\hat{m}_2 - m_2\|_{\mathcal{H}} = \|\hat{m}_{2,N} - m_2\|_{\mathcal{H}} = O_P(N^{-1/2})$. To this end we notice that by stationarity of $\{y_k^2\}$

$$\begin{aligned}E\|\hat{m}_2 - m_2\|_{\mathcal{H}}^2 &= E\left\|\frac{1}{N}\sum_{k=1}^{N}(y_k^2 - m_2)\right\|_{\mathcal{H}}^2 \\ &= \frac{1}{N}\sum_{k=-(N-1)}^{N-1}\left(1 - \frac{|k|}{N}\right)E\left\langle y_0^2 - m_2, y_k^2 - m_2\right\rangle \\ &\le \frac{1}{N}\left(E\left\|y_0^2 - m_2\right\|_{\mathcal{H}} + 2\sum_{k=1}^{\infty}\left|E\left\langle y_0^2 - m_2, y_k^2 - m_2\right\rangle\right|\right).\end{aligned}$$

By construction y_0^2 and the approximation y_{kk}^2 are independent. Repeated application of the Cauchy-Schwarz inequality together with Assumption 6.1 (a) yield that

$$\begin{aligned}\left|E\left\langle y_0^2 - m_2, y_k^2 - m_2\right\rangle\right| &= \left|E\left\langle y_0^2 - m_2, y_k^2 - y_{kk}^2\right\rangle\right| \\ &\le \left(E\left\|y_0^2 - m_2\right\|_{\mathcal{H}}^2\right)^{1/2}\left(E\left\|y_k^2 - y_{kk}^2\right\|_{\mathcal{H}}^2\right)^{1/2} \\ &= \left(E\left\|\sigma_0^2 - m_2\right\|_{\mathcal{H}}^2\right)^{1/2}\left(E\left\|\sigma_k^2 - \sigma_{kk}^2\right\|_{\mathcal{H}}^2\right)^{1/2}.\end{aligned}$$

Combining these estimates with Theorem 6.2 shows that $E\left\|\hat{m}_2 - m_2\right\|_{\mathcal{H}} = O(N^{-1/2})$. □

6.3.1 Estimation of β

We now turn to the estimation of the autoregressive operator β in the ARH(1) model (6.16). It is instructive to focus first on the univariate case $Z_n = \beta Z_{n-1} + \nu_n$, in which all quantities are scalars. We assume $E\nu_n = 0$ which implies $EZ_n = 0$. We also assume that $|\beta| < 1$, so that there is a stationary solution such that ν_n is uncorrelated with Z_{n-1}. Then, multiplying the AR(1) equation by Z_{n-1} and taking the expectation, we obtain $\gamma_1 = \beta\gamma_0$, where $\gamma_k = E[Z_n Z_{n+k}] = \mathrm{cov}(Z_n, Z_{n+k})$. The autocovariances γ_k are estimated in the usual way by the sample autocovariances

$$\hat{\gamma}_k = \frac{1}{N}\sum_{j=1}^{N-k} Z_j Z_{j+k},$$

so the usual estimator of β is $\hat{\beta} = \hat{\gamma}_1/\hat{\gamma}_0$. This is the so-called *Yule-Walker estimator* which is optimal in many ways, see Chapter 8 of Brockwell and Davis (1991).

In the functional setup we will replace condition $|\beta| < 1$ with $\|\beta\|_{\mathcal{S}} < 1$. Notice that this condition is guaranteed by Assumption 6.1 and that it will imply the existence of a weakly stationary solution of (6.16) of the form

$$Z_n = \sum_{j\geq 0} \beta^j(\nu_{n-j}),$$

where β^j is the j-times iteration of the operator β and β^0 is the identity mapping. The estimator for the operator β obtained in Bosq (2000) is formally analogue to the scalar case. We need instead of γ_0 and γ_1 the covariance operator

$$C_0(\cdot) = E\left[\langle Z_1, \cdot\rangle Z_1\right]$$

and the cross-covariance operator

$$C_1(\cdot) = E\left[\langle Z_1, \cdot\rangle Z_2\right].$$

One can show by similar arguments as in the scalar case that

$$\beta = C_1 C^{-1}.$$

To get an explicit form let $\lambda_1 \geq \lambda_2 \geq \cdots$ be the eigenvalues of C and let $e_1, e_2, \ldots$ be the corresponding eigenfunctions, i.e. $C(e_i) = \lambda_i e_i$. We assume that e_j are normalized to satisfy $\|e_j\|_{\mathcal{H}} = 1$. Then $\{e_j\}$ forms an orthonormal basis (ONB) of $\mathcal{H}$ and we obtain the following spectral decomposition of the operator C:

$$C(y) = \sum_{j\geq 1} \lambda_j \langle e_j, y\rangle e_j. \tag{6.19}$$

From (6.19) we get formally that

$$C^{-1}(y) = \sum_{j\geq 1} \lambda_j^{-1} \langle e_j, y\rangle e_j, \tag{6.20}$$

and hence

$$\begin{aligned}\beta(y) = C_1 C^{-1}(y) &= E\left(\left\langle Z_1, \sum_{j\geq 1} \lambda_j^{-1}\langle e_j, y\rangle e_j\right\rangle Z_2\right)\\ &= \sum_{j\geq 1} \lambda_j^{-1}\langle e_j, y\rangle E\left(\langle Z_1, e_j\rangle Z_2\right). \end{aligned}\tag{6.21}$$

Using $Z_2 = \sum_{i\geq 1}\langle Z_2, e_i\rangle e_i$ we obtain that the corresponding kernel is

$$\beta(t,s) = \sum_{i,j\geq 1} \lambda_j^{-1} E\left(\langle Z_1, e_j\rangle\langle Z_2, e_i\rangle\right) e_j(s) e_i(t). \tag{6.22}$$

If $\lambda_j = 0$ for all $j > p \geq 1$, then the covariance operator is finite rank and we can replace (6.19) and (6.20) by finite expansions with the sum going from 1 to p. In this case, all our mathematical operations so far are well justified. However, when all $\lambda_j > 0$ then we need to be aware that C^{-1} is not bounded on $\mathcal{H}$. To see this note that $\lambda_j \to 0$ if $j \to \infty$ (this follows from the fact that C is a Hilbert-Schmidt operator). Consequently, $\|C^{-1}(e_j)\|_{\mathcal{H}} = \lambda_j^{-1} \to \infty$ for $j \to \infty$. It can be easily seen that this operator is bounded only on

$$D = \left\{ y \in \mathcal{H} : \sum_{j \geq 1} \frac{\langle e_j, y \rangle^2}{\lambda_j^2} < \infty \right\}.$$

Nevertheless, we can show that the representation (6.21) holds for all $y \in \mathcal{H}$ by using a direct expansion of $\beta(t,s)$. Since the eigenfunctions $\{e_k,\, k \geq 1\}$ of C form an ONB of $\mathcal{H}$ it follows that $\{e_k \otimes e_\ell,\, k, \ell \geq 1\}$ ($e_k \otimes e_\ell = \{e_k(s)e_\ell(t),\, (s,t) \in [0,1]^2\}$) forms an ONB of $L^2([0,1]^2) = \mathcal{H} \otimes \mathcal{H}$. This is again a Hilbert space with inner product

$$\langle x, y \rangle_{\mathcal{H} \otimes \mathcal{H}} = \int \int x(t,s) y(t,s)\, dt\, ds.$$

Note that $\|\beta\|_{\mathcal{H} \otimes \mathcal{H}} = \|\beta\|_{\mathcal{S}} < \infty$ and hence the kernel function $\beta \in \mathcal{H} \otimes \mathcal{H}$. (Be aware, that for the sake of a lighter notation we don't distinguish between kernel and operator β.) Consequently $\beta(t,s)$ has the representation

$$\beta = \sum_{k,\ell \geq 1} \beta_{k,\ell} e_k \otimes e_\ell.$$

As we can write

$$Z_{n+1} = \sum_{k,\ell \geq 1} \beta_{k,\ell} \langle Z_n, e_k \rangle e_\ell + v_{n+1}$$

it follows that

$$\langle Z_{n+1}, e_i \rangle \langle Z_n, e_j \rangle = \sum_{k \geq 1} \beta_{k,i} \langle Z_n, e_k \rangle \langle Z_n, e_j \rangle + \langle \nu_{n+1}, e_i \rangle \langle Z_n, e_j \rangle$$

and by taking expectations on both sides of the above equation that

$$E\langle Z_2, e_i \rangle \langle Z_1, e_j \rangle = \sum_{k \geq 1} \beta_{k,i} \langle C(e_k), e_j \rangle = \beta_{j,i} \lambda_j.$$

Here we used the fact that $\{\nu_k\}$ is weak white noise. It implies that $E\langle B(v_k), x \rangle \langle v_\ell, y \rangle$ is zero for any bounded operator B and all $x, y \in \mathcal{H}$ and all $k \neq \ell$. Hence the expansion $Z_k = \sum_{j \geq 0} \beta^j(\nu_{k-j})$ provides $E\langle \nu_{n+1}, e_i \rangle \langle Z_n, e_j \rangle = 0$. This shows again (6.22).

We would like to obtain now an estimator for β by using a finite sample version of the above relations. To this end we set

$$\hat{C}(y) = \frac{1}{N}\sum_{k=1}^{N}\langle Z_k, y\rangle Z_k \quad \text{and} \quad \hat{C}_1(y) = \frac{1}{N}\sum_{k=1}^{N-1}\langle Z_k, y\rangle Z_{k+1}, \quad y \in \mathcal{H}.$$

The estimator in Bosq (2000) and the estimator we also propose here is of the form

$$\hat{\beta}(y; K) = \pi_K \hat{C}_1 \widehat{C^{-1}}(y; K),$$

where

$$\widehat{C^{-1}}(y; K) = \sum_{j=1}^{K} \hat{\lambda}_j^{-1}\langle \hat{e}_j, y\rangle \hat{e}_j, \tag{6.23}$$

$(\hat{\lambda}_j, \hat{e}_j)$ are the eigenvalues (in descending order) and the corresponding eigenfunctions of $\hat{C}$ and p_K is the orthogonal projection onto the subspace $\mathrm{span}(\hat{e}_1, \ldots, \hat{e}_K)$. We notice that this estimator is not depending on the sign of the $\hat{e}_j$'s. The corresponding kernel is given as

$$\hat{\beta}(t, s; K) = \frac{1}{N-1}\sum_{k=1}^{N-1}\sum_{j=1}^{K}\sum_{i=1}^{K}\hat{\lambda}_j^{-1}\langle Z_k, \hat{e}_j\rangle\langle Z_{k+1}, \hat{e}_i\rangle \hat{e}_j(s)\hat{e}_i(t), \tag{6.24}$$

and the signs of the $\hat{e}_j$ cancel out. In practice eigenvalues and eigenfunctions of an empirical covariance operator can be conveniently computed with the package `fda` for the statistical software `R`. The estimator (6.24) is the empirical version of the finite expansion

$$\beta(t, s; K) = \sum_{i=1}^{K}\sum_{j=1}^{K}\lambda_j^{-1}E\left(\langle Z_1, e_j\rangle\langle Z_2, e_i\rangle\right) e_j(s)e_i(t)$$

of (6.22).

If the innovations $\{\nu_k\}$ are i.i.d. Bosq (2000) proves under some technical conditions consistency of the estimator (6.24) when $K = K(N)$:

$$\|\beta - \hat{\beta}(K(N))\|_{\mathcal{L}} = o_P(1) \quad \text{as } N \to \infty.$$

The choice of $K(N)$ depends on the decay rate of the eigenvalues, which is not known in practice. Empirical results (Didericksen et al., 2010) show that in the finite sample case $K = 2, 3, 4$ provides best results. The reason why choosing small K is often favorable is due to a bias variance trade off. Note that the eigenvalues occur reciprocal in the estimator $\hat{\beta}$ and thus larger K accounts for larger instability if the eigenvalues are close to zero. A practical approach is to chose K the largest integer for which $\hat{\lambda}_K/\hat{\lambda}_1 \geq \gamma$, where γ is some threshold.

Theorem 6.4. *Fix some $K \geq 1$. Assume that the $K+1$ largest eigenvalues of the covariance operator C of Z_k satisfy $\lambda_1 > \lambda_2 > \ldots > \lambda_{K+1} > 0$. Let $\beta(K)$ and $\hat{\beta}(K)$ be the operators belonging to the kernel functions $\beta(t,s;K)$ and $\hat{\beta}(t,s;K)$, respectively. Let Assumption 6.1 hold with condition (b) strengthened to $\alpha = 4$. Then we have*

$$\|\beta(K) - \hat{\beta}(K)\|_{\mathcal{S}} = O_P\left(N^{-1/2}\right) \quad \text{as } N \to \infty.$$

In Theorem (6.4) N obviously denotes the sample size which is suppressed in the notation. The proof of the theorem is given in Section 6.5. Our conditions imply that $E\|Z_k\|^4 < \infty$. This assumption is probably more stringent than necessary and a relaxation would be desirable. Note however, that finite 4th moments are required in Bosq (2000) even for i.i.d. $\{\nu_k\}$.

6.3.2 Simulation study

In this section we demonstrate the capabilities of our estimators for $\beta(t,s)$ and $\delta(t)$ on simulated data. We proceed as follows: We will choose a simple $\beta(t,s)$ and $\delta(t)$, simulate several days of observations using these parameters, and then use the estimation procedure given in Section 6.3.1 to obtain $\hat{\beta}(t,s;2)$ and $\hat{\delta}(t;2)$ from (6.24) and (6.17) respectively.

We will use $\beta(t,s) = 16s(1-s)t(1-t)$ and $\delta(t) = 0.01$ for our simulations. Now that we have chosen $\beta(t,s)$ and $\delta(t)$ we can simulate data according to (6.1) and (6.2). We will use $\varepsilon_i(t) = B_i(t) + N_i\sqrt{1-t(1-t)}$ for the error term, where $B_i(t)$ are iid standard Brownian bridges and N_i are iid standard normals. Note that this gives $E(\varepsilon^2(t)) = 1$ for all t, which is assumed by our estimation procedure. After simulating N days of data we compute $\hat{\beta}(t,s;2)$ and $\hat{\delta}(t;2)$. Figures 6.1, 6.2, and 6.3 show the estimates when $N = 30$, $N = 300$, and $N = 3000$, respectively. We see from these plots that the estimators described in Section 6.3.1 accurately estimate the parameters, $\beta(t,s)$ and $\delta(t)$, when the sample size is sufficiently large. Note that each plot of $\hat{\delta}(t;2)$ has the true $\delta(t)$ superimposed. A plot of the true $\beta(t,s)$ is given in Figure 6.4.

6.4 An example

In this section we show an example illustrating that our model captures the basic features of intraday returns. Let $P_k(t)$ denote the price of a stock on day k at time t. Then $y_k(t)$ can be viewed as the log-returns of the stock, $y_k(t) = \log P_k(t) - \log P_k(t-h)$, during period h (Cyree et al., 2004), where h is typically 1, 5, or 15 minutes. We will use $h = 5$ for 5-minute returns. The volatility of the stock is then represented by $\sigma_k^2(t) = \mathrm{Var}(y_k(t)|\mathcal{F}_{k-1})$.

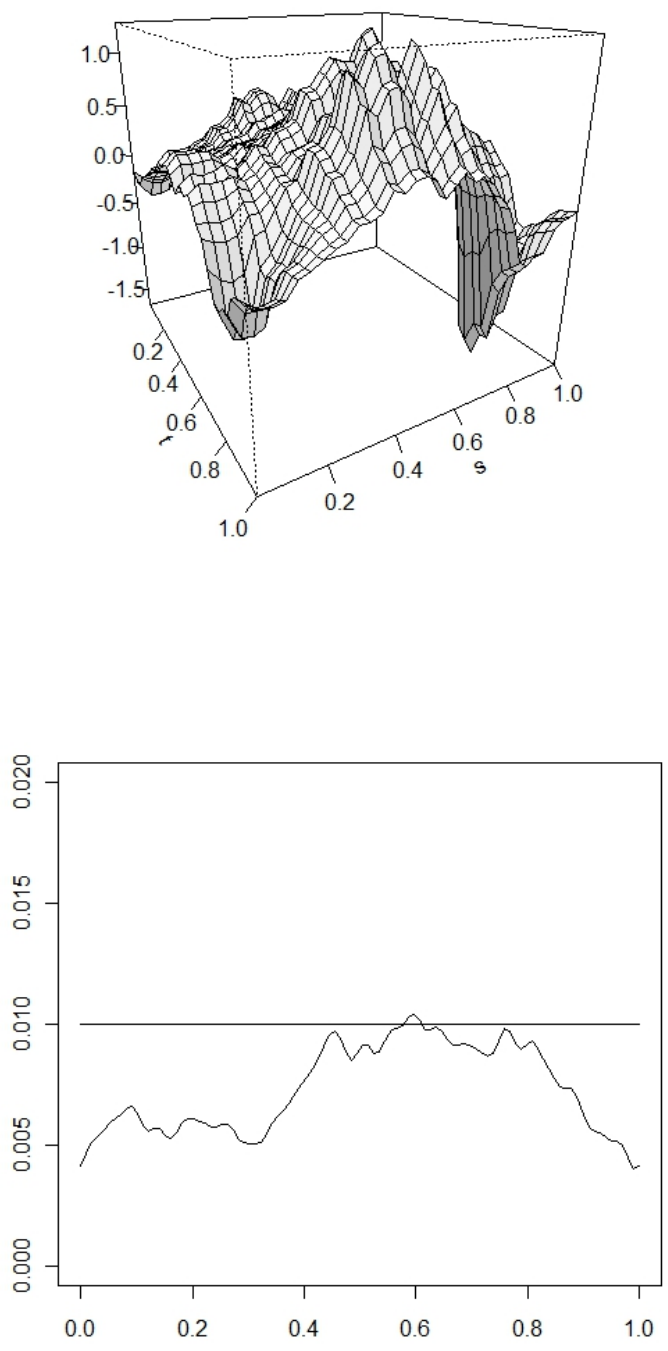
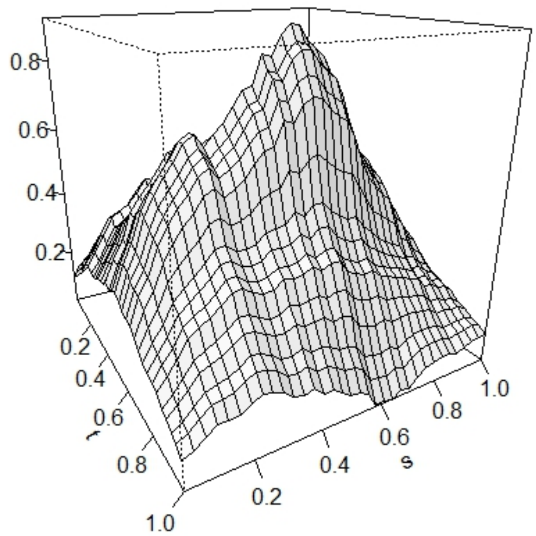


Figure 6.1. Using a sample of size $N = 30$, we obtain $\hat{\beta}(t, s; 2)$ on the top and $\hat{\delta}(t; 2)$ with $\delta(t) = .01$ superimposed on the bottom.



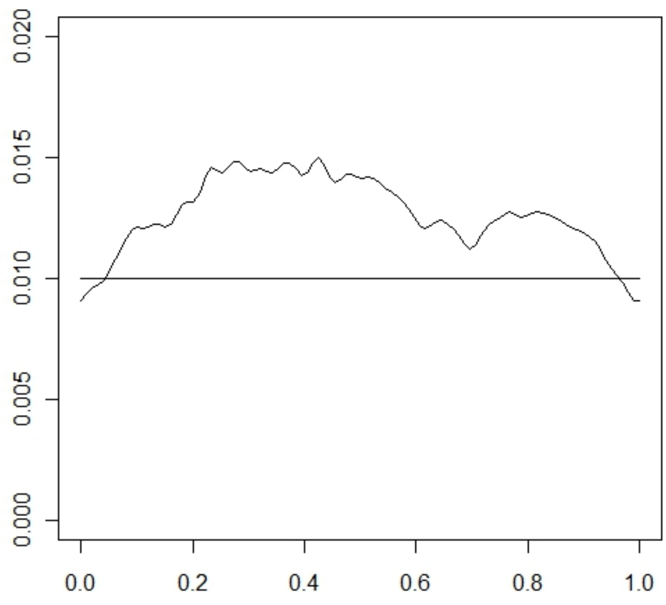


Figure 6.2. Using a sample of size $N = 300$, we obtain $\hat{\beta}(t, s; 2)$ on the top and $\hat{\delta}(t; 2)$ with $\delta(t) = .01$ superimposed on the bottom.

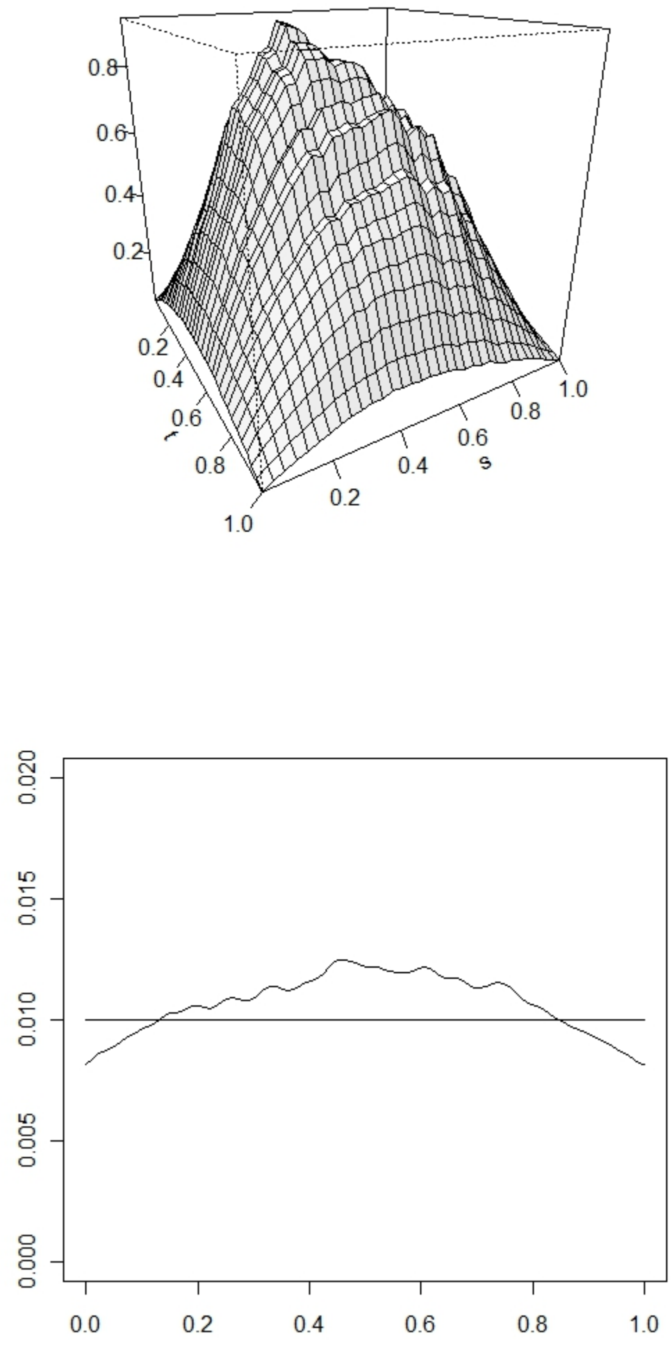


Figure 6.3. Using a sample of size $N = 3000$, we obtain $\hat{\beta}(t, s; 2)$ on the top and $\hat{\delta}(t; 2)$ with $\delta(t) = .01$ superimposed on the bottom.

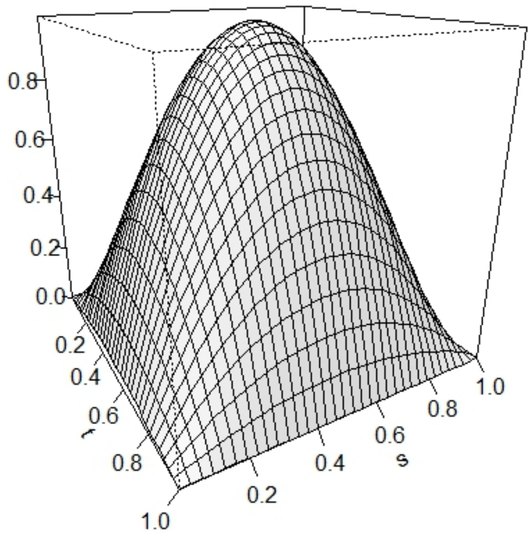


Figure 6.4. $\beta(t,s) = 16s(1-s)t(1-t)$

The first step to simulating the intraday returns is to estimate the parameters, $\delta(t)$ and $\beta(t,s)$, as outlined in Section 6.3.1. These parameters were estimated for the S&P 100 index based on data from April 1, 1997 to March 30, 2007. The estimated functions, $\hat{\beta}(t,s;2)$ and $\hat{\delta}(t;2)$, are shown in Figure 6.5.

Notice in Figure 6.5 that $\hat{\beta}(t,s;2)$ and $\hat{\delta}(t;2)$ are somewhat larger when t is close to 0 or 1. According to (6.2) this suggests that the volatility, $\sigma_k^2(t)$, tends to be larger at the beginning and end of each trading day. Higher volatilities at the beginning and the end of the trading day have been observed by several authors (Gau, 2005; Evans and Speight, 2010). This phenomenon is consistent with our observed log-return data based on the S&P 100 index and is captured by our model.

Having estimated the parameters, $\delta(t)$ and $\beta(t,s)$, we can now simulate several days of observations according to (6.1) and (6.2). We will use $\varepsilon_i(t) = 2^{-200t}\sqrt{\log(2)}W_i(2^{400t}/\log(2))$ for the error term, where $W_i(t)$ are iid standard Brownian motions. Note that this gives $E(\varepsilon^2(t)) = 1$ for all t, which is assumed by our estimation procedure.

We simulated 3 days of log-returns which we compare with the log-returns of the S&P 100 index. Figure 6.6 is a plot of the 5-minute returns on the S&P 100 index between April 11 and April 13, 2000. Figure 6.7 shows three consecutive days of simulated values for $y_k(t)$. The simulations show that our model empirically captures the main characteristics of financial data.

6.5 Proofs

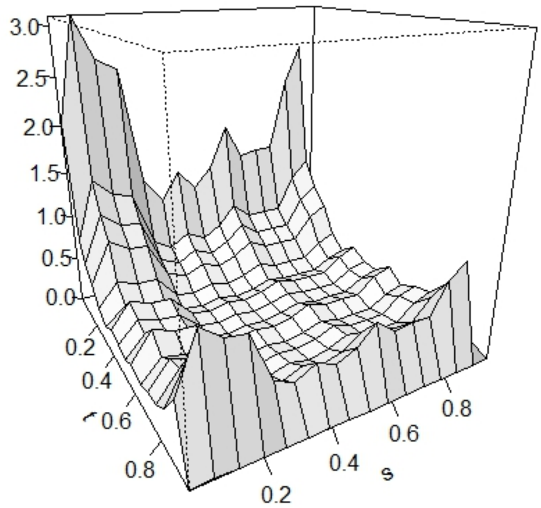
The proofs of Theorems 6.1 and 6.3 are based on general results for iterated random functions as those in Wu and Shao (2004); Diaconis and Freedman (1999). For the convenience of the reader we shall repeat here the main ideas of Wu and Shao (2004).

Let (S,ρ) be a complete, separable metric space. Let Θ be another metric space and let $M : \Theta \times S \to S$ be a measurable function. For a random element θ with values in Θ, an iterated random function system is defined via the random mappings $M_\theta(\cdot)$. More precisely it is assumed that

$$X_n = M_{\theta_n}(X_{n-1}), \quad n \in \mathbb{N}, \tag{6.25}$$

where $\{\theta_n,\ -\infty < n < \infty\}$ is an i.i.d. sequence with values in Θ. Thereby it is assumed that X_0 is independent of $\{\theta_n,\, n \geq 1\}$. For any $x \in S$ we define

$$S_n(x) = M_{\theta_n} \circ M_{\theta_{n-1}} \circ \cdots \circ M_{\theta_1}(x), \quad n = 1, 2, \ldots,$$



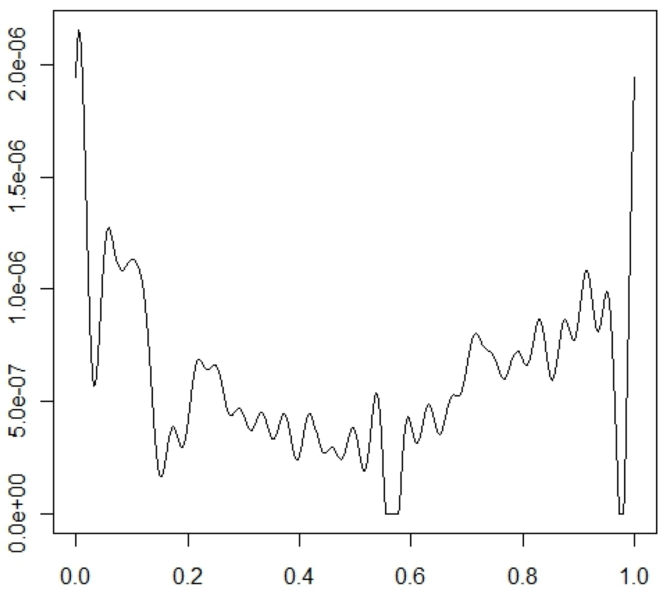


Figure 6.5. Estimated from the S&P 100 index: Top: $\hat{\beta}(t, s; 2)$. Bottom: $\hat{\delta}(t; 2)$.

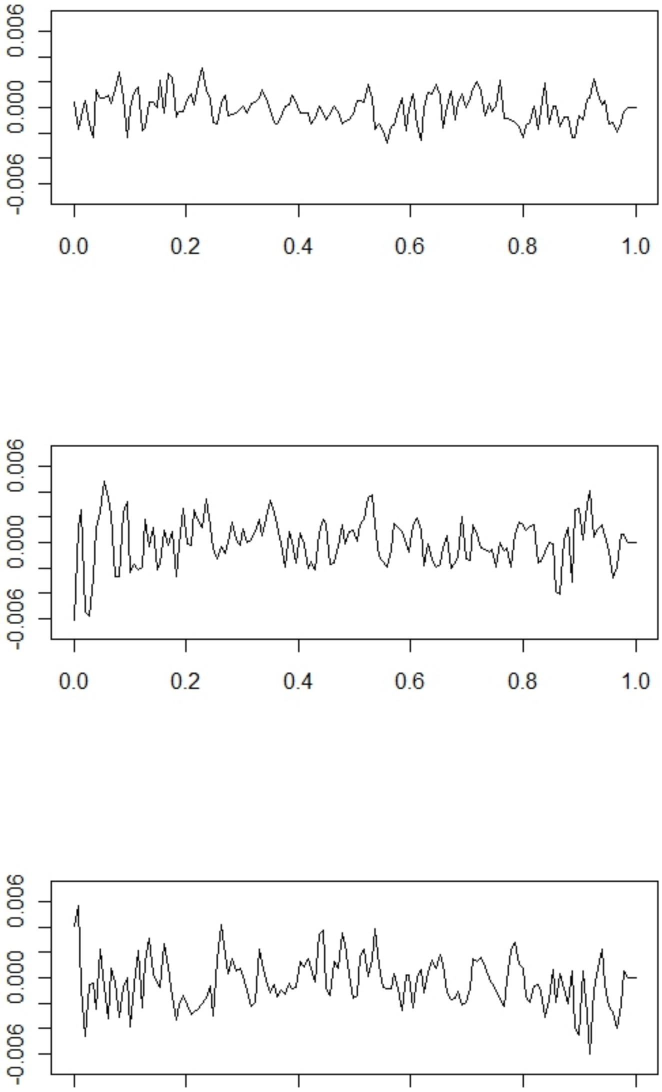


Figure 6.6. 5-minute log-returns for the S&P 100 index between April 11 and April 13, 2000.

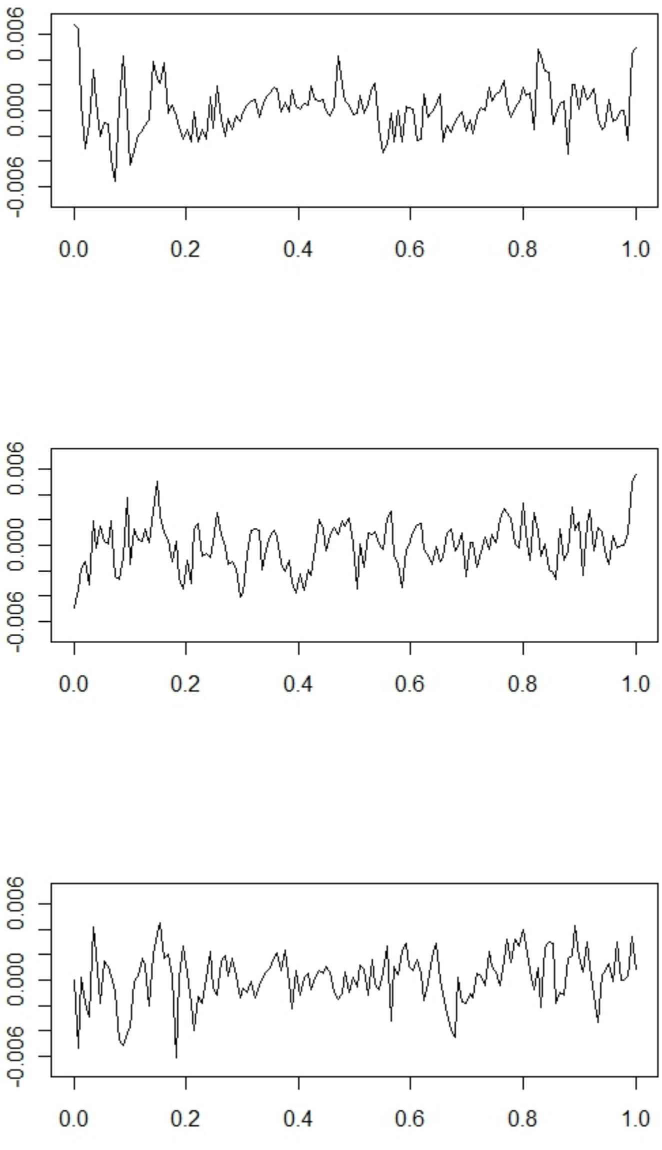


Figure 6.7. Three consecutive days of simulated values for $y_k(t)$.

where $\circ$ denotes the composition of functions. We also introduce the backward version of S_n, which is given by

$$Z_n(x) = M_{\theta_{-1}} \circ M_{\theta_{-2}} \circ \cdots \circ M_{\theta_{-n}}(x), \quad x \in S,\ n = 1, 2, \ldots .$$

The following theorem is a slight modification of Theorem 2 of Wu and Shao (2004), so that it is immediately applicable for our purposes.

Theorem 6.5. *(Wu and Shao, 2004) Assume that*

(A) *there are* $y_0 \in S$ *and* $\alpha > 0$ *such that* $E\{\rho(y_0, M_{\theta_0}(y_0))\}^\alpha < \infty$ *and*

(B) *there are* $x_0 \in S$, $\alpha > 0$, $0 < r_1 = r_1(\alpha) < 1$ *and* $c = c(\alpha) < \infty$ *such that*

$$E\{\rho(S_n(x), S_n(x_0))\}^\alpha \leq c r_1^n \{\rho(x, x_0)\}^\alpha$$

for all $x \in S$ *and* $n \in \mathbb{N}$. *Then for all* $x \in S$ *we have* $Z_n(x)$ *converges almost surely to some* Z_∞ *which is independent of* x. *Furthermore* $Z_\infty = g(\theta_0, \theta_{-1}, \ldots)$ *and*

$$E\{\rho(Z_n(x), Z_\infty)\}^\alpha \leq c_1 r^n,$$

where $c_1 = c_1(x, x_0, y_0, \alpha) < \infty$ *and* $0 < r = r(\alpha) < 1$. *Moreover, the process* $X_n = g(\theta_n, \theta_{n-1}, \ldots)$ *is a stationary solution of* (6.25). *Finally, if we let* $X_0^* = f(\theta'_0, \theta'_{-1}, \ldots)$ *where* $\{\theta'_n\}$ *is an independent copy of* $\{\theta_n\}$, *then*

$$E\{\rho(S_n(X_0^*), S_n(X_0))\}^\alpha \leq c_2 r_2^n,$$

with some $0 < r_2 = r_2(\alpha) < 1$ *and* $c_2 = c_2(\alpha) > 0$.

Proof of Theorem 6.1. We need to show that the conditions of Theorem 6.5 are satisfied when the underlying space is $\mathcal{H}$ with metric $\|\cdot\|_{\mathcal{H}}$ and

$$M_{\theta_n}(x)(t) = \delta(t) + \int \beta(t,s)\varepsilon_{n-1}^2(s)x(s)\,ds.$$

To demonstrate *(A)* of Theorem 6.5 we use $y_0(t) = 0$, $0 \leq t \leq 1$, and get

$$\|y_0 - M_{\theta_0^2}(y_0)\|_{\mathcal{H}}^2 = \int \delta^2(t)\,dt < \infty,$$

by assumption. Since for any $x, x_0 \in \mathcal{H}$ we have

$$\begin{aligned}
\|S_n(x) - S_n(x_0)\|_{\mathcal{H}} &= \|M_{\theta_n^2}(S_{n-1}(x)) - M_{\theta_n^2}(S_{n-1}(x_0))\|_{\mathcal{H}} \\
&= \left(\int \left(\int \beta(t,s)\{S_{n-1}(x)(s) - S_{n-1}(x_0)(s)\}\varepsilon_{n-1}^2(s)\,ds \right)^2 dt \right)^{1/2} \\
&\leq \left(\int \left\{ \int \beta^2(t,s)\varepsilon_{n-1}^4(s)\,ds \right\} \int \left\{ S_{n-1}(x)(s) - S_{n-1}(x_0)(s) \right\}^2 ds\,dt \right)^{1/2} \\
&= K(\varepsilon_{n-1}^2)\,\|S_{n-1}(x) - S_{n-1}(x_0)\|_{\mathcal{H}},
\end{aligned}$$

by the Cauchy-Schwarz inequality. Repeating the arguments above, we conclude

$$\big\|S_n(x) - S_n(x_0)\big\|_{\mathcal{H}} \leq \big\|x - x_0\big\|_{\mathcal{H}} \prod_{i=0}^{n-1} K(\varepsilon_i^2).$$

Taking expectations on both sides and using the independence of the ε_i proves *(B)*. $\square$

Theorem 6.2 is a simple corollary to Theorem 6.1 and Theorem 6.5. Theorem 6.3 can be proven along the same lines of argumentation and the proof is omitted.

Proof of Propositions 6.1 and 6.2. First we establish (6.7). We follow the proof of Theorem 6.1. Since $E\big\{\|\sigma_0^2\|_{\mathcal{H}}\big\}^\alpha = E\big\{\|Z_\infty\|_{\mathcal{H}}\big\}^\alpha$, according to the construction in the proof of Theorem 6.1 we have

$$\begin{aligned}
E\big\{\|\sigma_0^2\|_{\mathcal{H}}\big\}^\alpha &= E\big\{\|Z_\infty\|_{\mathcal{H}}\big\}^\alpha \\
&\leq E\big\{\|Z_1(0)\|_{\mathcal{H}} + \|Z_1(0) - Z_\infty(0)\|_{\mathcal{H}}\big\}^\alpha \\
&\leq 2^\alpha \Big\{E\big\{\|Z_1(0)\|_{\mathcal{H}}\big\}^\alpha + E\big\{\|Z_1(0) - Z_\infty(0)\|_{\mathcal{H}}\big\}^\alpha\Big\},
\end{aligned}$$

where 0 denotes the "zero function" on $[0,1]$. According the proof of Theorem 6.1 and Theorem 6.5 the term $E\big\{\|Z_1(0) - Z_\infty(0)\|_{\mathcal{H}}\big\}^\alpha < \infty$. Furthermore, the term $E\big\{\|Z_1(0)\|_{\mathcal{H}}\big\}^\alpha = \Big(\int \delta^2(t)\,dt\Big)^{\frac{\alpha}{2}} < \infty$. To show (6.10), we note that

$$\begin{aligned}
E\big\{\|y_0\|_{\mathcal{H}}\big\}^\alpha &= E\bigg[\int y_0^2(t)\,dt\bigg]^{\frac{\alpha}{2}} \\
&= E\bigg[\int \varepsilon_0^2(t)\sigma_0^2(t)\,dt\bigg]^{\frac{\alpha}{2}} \\
&\leq E\big\{\|\varepsilon_0\|_\infty\big\}^\alpha E\big\{\|\sigma_0\|_{\mathcal{H}}\big\}^\alpha,
\end{aligned}$$

since ε_0 and σ_0 are independent processes. Proposition 6.1 is proven.

The proof of Proposition 6.2 only requires minor modifications and is therefore omitted. $\square$

Proof of Proposition 6.3. Using recursion 6.1 we have

$$\begin{aligned}
\omega(y_0, h) &= \sup_{0\leq t\leq 1-h}\,\sup_{0\leq s\leq h} |\varepsilon_0(t+s)\sigma_0(t+s) - \varepsilon_0(t)\sigma_0(t)| \\
&\leq \sup_{0\leq t\leq 1-h}\,\sup_{0\leq s\leq h} \big\{|\varepsilon_0(t+s)||\sigma_0(t+s) - \sigma_0(t)| + |\sigma_0(t)||\varepsilon_0(t+s) - \varepsilon_0(t)|\big\} \\
&\leq \|\varepsilon_0\|_\infty \omega(\sigma_0, h) + \|\sigma_0\|_\infty \omega(\varepsilon_0, h).
\end{aligned}$$

The independence of ε_0 and σ_0 yields

$$E\big\{\|\sigma_0\|_\infty \omega(\varepsilon_0, h)\big\}^p = E\big\{\|\sigma_0\|_\infty\big\}^p E\big\{\omega(\varepsilon_0, h)\big\}^p.$$

Proposition 6.2 gives $E\{\|\sigma_0^2\|_\infty\}^p < \infty$. This implies that $E\{\|\sigma_0\|_\infty\}^p < \infty$ and therefore

$$\lim_{h\to 0} E\{\|\sigma_0\|_\infty \omega(\varepsilon_0, h)\}^p = 0.$$

The identity $|\sqrt{a} - \sqrt{b}| \le \sqrt{|a-b|}$, $a, b \ge 0$, implies

$$\begin{aligned}\omega^p(\sigma_0, h) &= \sup_{0\le t\le 1-h} \sup_{0\le s\le h} |\sigma_0(t+s) - \sigma_0(t)|^p \\ &\le \left(\sup_{0\le t\le 1-h} \sup_{0\le s\le h} |\sigma_0^2(t+s) - \sigma_0^2(t)| \right)^{\frac{p}{2}}.\end{aligned}$$

Recursion (6.2) gives

$$\begin{aligned}|\sigma_0^2(t+s) - \sigma_0^2(t)| &\le |\delta(t+s) - \delta(t)| + \left| \int (\beta(t+s, r) - \beta(t, r)) y_{-1}^2(r) dr \right| \\ &\le \omega(\delta, h) + \sup_{0\le t\le 1-h} \sup_{0\le s\le h} \sup_{0\le r\le 1} |\beta(t+s, r) - \beta(t, r)| \times \int y_{-1}^2(r) dr.\end{aligned}$$

Hence

$$\begin{aligned}E\left(\omega^p(\sigma_0, h)\right) &= E\{ \sup_{0\le t\le 1-h} \sup_{0\le s\le h} |\sigma_0^2(t+s) - \sigma_0^2(t)| \}^{\frac{p}{2}} \\ &\le 2^{\frac{p}{2}} \Big\{ \big[\sup_{0\le t\le 1-h} \sup_{0\le s\le h} \sup_{0\le r\le 1} |\beta(t+s, r) - \beta(t, r)| \big]^{\frac{p}{2}} \times E\{\|y_0\|_\infty\}^p + \big[\omega(\delta, h)\big]^{\frac{p}{2}} \Big\}.\end{aligned}$$

Proposition 6.2 yields that $E\{\|y_0\|_\infty\}^p < \infty$ and $E\{\|\sigma_0\|_\infty\}^p < \infty$. So by the independence of the processes ε_0 and σ_0 we conclude

$$\lim_{h\to 0} \{\|\varepsilon_0\|_\infty \omega(\sigma_0, h)\}^p = 0,$$

completing the proof of Proposition 6.3. □

Proof of Theorem 6.4. Under our assumptions it follows from Theorem 6.2 that for any $m \ge 1$

$$E\|Z_k - Z_{km}\|_{\mathcal{H}}^4 \le \text{const} \times r^m$$

where $r \in (0, 1)$ and Z_{km} are the m-dependent approximations of Z_k (constructed by using σ_{km}^2 instead of σ_k^2 in the definition of Z_k). This shows that the notion of L^4-m-

approximability suggested in Hörmann and Kokoszka (2010) applies to the sequence $\{Z_k\}$. As consequence we have with $\hat{c}_i = \operatorname{sign}\langle \hat{e}_i, e_i \rangle$ that

$$
\begin{aligned}
&(a) \quad \max_{1 \le i \le K} E\|\hat{c}_i \hat{e}_i - e_i\|_{\mathcal{H}}^2 = O\left(N^{-1}\right); \\
&(b) \quad \max_{1 \le i \le K} E|\hat{\lambda}_i - \lambda_i|^2 = O\left(N^{-1}\right)
\end{aligned}
$$

and therefore that

$$
\begin{aligned}
&(a') \quad \max_{1 \le i \le K} \|\hat{c}_i \hat{e}_i - e_i\|_{\mathcal{H}} = O_P\left(N^{-1/2}\right); \\
&(b') \quad \max_{1 \le i \le K} |\hat{\lambda}_i - \lambda_i| = O_P\left(N^{-1/2}\right)
\end{aligned}
$$

See Theorem 3.2 of Hörmann and Kokoszka (2010). The random sign $\hat{c}_i$ (which we cannot observe) accounts for the fact that e_i can be only uniquely identified up to its sign. As our estimator $\hat{\beta}(K)$ doesn't depend on the signs of the $\hat{e}_i$, this poses no problem. We define

$$\sigma_{i,j} = E\langle Z_1, e_i \rangle \langle Z_2, e_j \rangle$$

and let

$$\hat{\sigma}_{i,j} = \frac{1}{N-1} \sum_{k=1}^{N-1} \langle Z_k, \hat{e}_i \rangle \langle Z_{k+1}, \hat{e}_j \rangle$$

be the empirical counterpart. Then we have

$$
\begin{aligned}
E|\sigma_{i,j} - \hat{c}_i \hat{c}_j \hat{\sigma}_{i,j}| \le E &\left| \frac{1}{N-1} \sum_{k=1}^{N-1} \left(\langle Z_k, e_i \rangle \langle Z_{k+1}, e_j \rangle - E\langle Z_k, e_i \rangle \langle Z_{k+1}, e_j \rangle \right) \right| \\
&+ E\left| \frac{1}{N-1} \sum_{k=1}^{N-1} \left(\langle Z_k, \hat{c}_i \hat{e}_i \rangle \langle Z_{k+1}, \hat{c}_j \hat{e}_j \rangle - \langle Z_k, e_i \rangle \langle Z_{k+1}, e_j \rangle \right) \right| \\
&=: T_1(i, j; N) + T_2(i, j; N).
\end{aligned}
$$

The processes $\mathcal{Z}_k = \mathcal{Z}_k(i, j) = \langle Z_k, e_i \rangle \langle Z_{k+1}, e_j \rangle$ are strictly stationary for every choice of i and j and we can again define the approximations $\mathcal{Z}_{km}$ in the spirit of Section 6.2. We have by independence of $\mathcal{Z}_0$ and $\mathcal{Z}_{kk}$

$$
\begin{aligned}
\sum_{h \ge 0} |\operatorname{Cov}(\mathcal{Z}_0, \mathcal{Z}_h)| &\le E\mathcal{Z}_0^2 + \left(E\mathcal{Z}_0^2\right)^{1/2} \times \sum_{h \ge 1} \left(E(\mathcal{Z}_h - \mathcal{Z}_{hh})^2\right)^{1/2} \\
&\le E\|Z_0\|_{\mathcal{H}}^4 + E\|Z_0\|_{\mathcal{H}}^2 \times \sum_{h \ge 1} \left(E(\mathcal{Z}_h - \mathcal{Z}_{hh})^2\right)^{1/2}.
\end{aligned}
$$

Further we have by repeated application of the Cauchy-Schwarz inequality that

$$\begin{aligned}
E(\mathcal{Z}_h - \mathcal{Z}_{hh})^2 &= E\left[\langle Z_h, e_i\rangle\langle Z_{h+1}, e_j\rangle - \langle Z_{hh}, e_i\rangle\langle Z_{h+1,h}, e_j\rangle\right]^2 \\
&\le 2\left\{E\left[\langle Z_h - Z_{hh}, e_i\rangle\langle Z_{h+1}, e_j\rangle\right]^2 + E\left[\langle Z_{hh}, e_i\rangle\langle Z_{h+1} - Z_{h+1,h}, e_j\rangle\right]^2\right\} \\
&\le 2\left\{E\langle Z_h - Z_{hh}, e_i\rangle^2 E\langle Z_h, e_j\rangle^2 + E\langle Z_{hh}, e_i\rangle^2 E\langle Z_{h+1} - Z_{h+1,h}, e_j\rangle^2\right\} \\
&\le 2E\|Z_0\|_{\mathcal{H}}^2\left\{E\|Z_h - Z_{hh}\|_{\mathcal{H}}^2 + E\|Z_{h+1} - Z_{h+1,h}\|_{\mathcal{H}}^2\right\} \\
&\le \text{const} \times r^h,
\end{aligned}$$

for some $r \in (0,1)$. This proves that the autocovariances of the process $\{\mathcal{Z}_k\}$ are absolutely summable. A well known result in time series analysis thus implies that

$$(N-1)\mathrm{Var}\left(\frac{1}{N-1}\sum_{k=1}^{N-1}\mathcal{Z}_k\right) \le 2\sum_{h\ge 0}|\mathrm{Cov}(\mathcal{Z}_0, \mathcal{Z}_h)| \le c_0, \quad \forall N \ge 2,$$

See the proof of Theorem 7.1.1 in Brockwell and Davis (1991), where, as we have shown, the constant c_0 is independent of the choice of i and j in the definition of $\mathcal{Z}_k$. Hence $\max_{1\le i,j\le K} T_1(i,j;N) = O_P\left(N^{-1/2}\right)$.

Using relation (a) above, one can show that also $\max_{1\le i,j\le K} T_2(i,j;N) = O_P\left(N^{-1/2}\right)$. We thus have

$$(c) \qquad \max_{1\le i,j\le K}|\sigma_{i,j} - \hat{c}_i\hat{c}_j\hat{\sigma}_{i,j}| = O_P\left(N^{-1/2}\right).$$

We have now the necessary tools to prove Theorem 6.4. By relations (a'), (b') and (c) we have that

$$\begin{aligned}
\left\|\beta(K) - \hat{\beta}(K)\right\|_{\mathcal{S}} &= \left\|\sum_{1\le i,j\le K}\left(\frac{\sigma_{j,i}}{\lambda_j}e_j \otimes e_i - \frac{\hat{\sigma}_{j,i}}{\hat{\lambda}_j}\hat{e}_j \otimes \hat{e}_i\right)\right\|_{\mathcal{S}} \\
&\le \sum_{1\le i,j\le K}\left\{\left|\frac{\sigma_{j,i}}{\lambda_j} - \frac{\hat{c}_j\hat{c}_i\hat{\sigma}_{j,i}}{\hat{\lambda}_j}\right| + \left|\frac{\sigma_{j,i}}{\hat{\lambda}_j}\right|\|e_j \otimes e_i - \hat{c}_j\hat{e}_j \otimes \hat{c}_i\hat{e}_i\|_{\mathcal{S}}\right\} \\
&\le K^2\left\{O_P\left(N^{-1/2}\right) + O_P(1)\max_{1\le i,j\le K}\|e_j \otimes e_i - \hat{c}_j\hat{e}_j \otimes \hat{c}_i\hat{e}_i\|_{\mathcal{S}}\right\}.
\end{aligned}$$

The proof follows from $\|e_j \otimes e_i - \hat{c}_j\hat{e}_j \otimes \hat{c}_i\hat{e}_i\|_{\mathcal{S}} \le \|\hat{c}_j\hat{e}_j - e_j\|_{\mathcal{H}} + \|\hat{c}_i\hat{e}_i - e_i\|_{\mathcal{H}}$. □

6.6 Bibliography

Angelidis, T. and S. Degiannakis (2008). Volatility Forecasting: Intra-day Versus Inter-day Models. *Journal of International Financial Markets, Institutions & Money 18*, 449–465.

Aue, A., S. Hörmann, L. Horváth, M. Hušková, and J. Steinebach (2011). Sequential Testing for the Stability of Portfolio Betas. *Econometric Theory 00*, 000–000.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2008). Designing Realised Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise. *Econometrica 76*, 1481–1536.

Barndorff-Nielsen, O. and N. Shephard (2004). Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics. *Econometrica 72*, 885–925.

Bollerslev, T. (1990). Modelling the Coherence in Short Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. *Review of Economics and Statistics 72*, 498–505. Reprinted in *ARCH: Selected Readings* (ed. R. F. Engle), Oxford University Press, 1995.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods*. Springer.

Cyree, K. B., M. D. Griffiths, and D. B. Winters (2004). An Empirical Examination of the Intraday Volatility in Euro-Dollar Rates. *The Quarterly Review of Economics and Finance 44*, 44–57.

Diaconis, P. and D. Freedman (1999). Iterated Random Functions. *Siam Review 41*, 45–76.

Didericksen, D., P. Kokoszka, and X. Zhang (2010). Empirical Properties of Forecasts with the Functional Autoregressive Model. Technical report, Utah State University.

Elezović, S. (2009). Functional Modelling of Volatility in the Swedish Limit Order Book. *Computational Statistics & Data Analysis 53*, 2107–2118.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica 50*, 987–1007.

Evans, K. and A. Speight (2010). Intraday Periodicity, Calendar and Announcement Effects in Euro Exchange Rate Volatility. *Research in International Business and Finance 24*, 82–101.

Fatum, R. and J. Pedersen (2009). Real-time Effects of Central Bank Intervention in the Euro Market. *Journal of International Economics 78*, 11–20.

Gau, Y.-F. (2005). Intraday Volatility in the Taipei FX Market. *Pacific–Basin Finance Journal 13*, 471–487.

Harrison, J., R. Pitbladdo, and S. Schaefer (1984). Continuous Prices in Frictionless Markets have Infinite Variation. *Journal of Business 57*, 353–365.

Hörmann, S. and P. Kokoszka (2010). Weakly Dependent Functional Data. *The Annals of Statistics 38*, 1845–1884.

Jacod, J., Y. Li, P. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure Noise in the Continuous Case: The Pre-averaging Approach. *Stochastic Processes and their Applications 119*, 2249–2276.

Jeantheau, T. (1998). Strong Consistency of Estimators for Multivariate ARCH Models. *Econometric Theory 14*, 70–86.

Ling, S. and M. McAleer (2002). Necessary and Sufficient Moment Conditions for the GARCH(p,q) and Asymmetric Power GARCH(p,q) Models. *Econometric Theory 18*, 722–729.

Nelson, D. B. (1990). Stationarity and Persistence in the GARCH$(1, 1)$ Model. *Econometric Theory 6*, 318–334.

Silvennoinen, A. and T. Teräsvirta (2009). Multivariate GARCH Models. In T. Andersen, R. Davis, J. Krei, and T. Mikosch (Eds.), *Handbook of Financial Time Series*. Springer.

Stout, W. F. (1974). *Almost Sure Convergence*. Academic Press.

Wu, W. and X. Shao (2004). Limit Theorems for Iterated Random Functions. *Journal of Applied Probability 41*, 425–436.

Zhang, S. R., J. M. Holt, A. P. van Eyken, M. McCready, C. Amory-Mazaudier, S. Fukao, and M. Sulzer (2005). Ionospheric Local Model and Climatology from Long-term Databases of Multiple Incoherent Scatter Radars. *Geophys. Res. Lett. 32*, doi:10.1029/2005GL023603.